TIME-CONSISTENT NO-ARBITRAGE
MODELS OF THE TERM STRUCTURE

Michael W. Brandt
Amir Yaron

Time-Consistent No-Arbitrage Models of the Term Structure
Michael W. Brandt and Amir Yaron
NBER Working Paper No. 9458
January 2003
JEL No. G0

ABSTRACT

We present an econometric procedure for calibrating no-arbitrage term structure models in a way that is time-consistent and robust to measurement errors. Typical no-arbitrage models are time-inconsistent because their parameters are assumed constant for pricing purposes despite the fact that the parameters change whenever the model is recalibrated. No-arbitrage models are also sensitive to measurement errors because they fit exactly each potentially contaminated bond price in the cross-section. We overcome both problems by evaluating bond prices using the joint dynamics of the factors and calibrated parameters and by locally averaging out the measurement errors. Our empirical application illustrates the trade-off between fitting as well as possible and overfitting the cross-section of bond prices due to measurement errors. After optimizing this trade-off, our approach fits almost exactly the cross-section of bond prices at each date and produces out-of-sample forecast errors that beat a random walk benchmark and are comparable to the results in the affine term structure literature. We find that non-linearities in the pricing kernel are important, lending support to quadratic term structure models.

Michael W. Brandt
The Wharton School
University of Pennsylvania
2300 Steinberg-Deitrich Hall
Philadelphia, PA 19104-6367
and NBER
brandtm@wharton.upenn.edu

Amir Yaron
The Wharton School
University of Pennsylvania
2256 Steinberg-Deitrich Hall
Philadelphia, PA 19104-6367
and NBER
yaron@wharton.upenn.edu

1 Introduction

No-arbitrage models of the term structure are popular among practitioners because they provide a description of the yield curve that is consistent with the prices of all actively traded bonds on a given date and can therefore be used for pricing less liquid bonds and fixed income derivatives on the same date.[1] Although often categorized as a separate class of "models," no-arbitrage models have the same theoretical underpinnings as other term structure models. The difference is really an econometric issue. No-arbitrage models are calibrated to the cross-section of bond prices on a single date, while the more academic approach to estimating term structure models involves a historical panel of bond prices.

No-arbitrage models suffer from at least three problems. The first problem is conceptual. Calibrating the model on one date and then recalibrating it on the next, which implicitly allows the parameters of the model to change, is inconsistent with the theoretical structure of the model that assumes constant parameters (Dybvig, 1989). The econometric procedure is therefore *time-inconsistent* because the econometrician *pretends* that the parameters are constant when fitting the model *knowing* that the parameters will be different when the model is refitted next period. The second problem is econometric. Since the model is never confronted with historical data there is no guarantee that the dynamics of the factors driving the time-variation of the yield curve implied by the cross-sectional fit correspond to the actual dynamics of the factors through time. If the factor dynamics are misspecified, the model may admit arbitrage despite appearing arbitrage-free (Backus, Foresi, and Zin, 1998). Finally, from a more practical perspective, no-arbitrage models are extremely sensitive to measurement errors, market microstructure effects, and other data imperfections, because they fit exactly each potentially contaminated bond price in the cross-section.

We present an econometric procedure that overcomes these problems. We achieve time-consistency and well-specified dynamics in a few steps. First, we separate parts of the model that can be identified outside the bond market, namely exogenous factors driving the endowment side of the economy, from parts that are inherently related to asset prices, namely aggregate preferences. Assuming the endowment side of the economy does not change from one date to the next, we estimate the dynamics of the factors from historical data and impose these dynamics on the cross-sectional fit of the arbitrage-free model at each date. Second, we infer from the cross-section of bond prices at each date the preference parameters that make the model approximately arbitrage-free. Third, we estimate and explicitly incorporate the

---

[1]No-arbitrage models calibrated to bond prices (and in some cases also their volatilities) on a single date include Ho and Lee (1986), Black, Derman, and Toy (1990), Hull and White (1990,1993), Black and Karasinski (1991), Jamshidian (1991), and Heath, Jarrow, and Morton (1992).

dynamics of the preference parameters from one cross-sectional fit to the next into the bond pricing equation. Since the second and third step are intimately related (we need the bond pricing equation to infer the preference parameters and we need realizations of the preference parameters to estimate and incorporate their dynamics into the bond pricing equation), we implement these two steps in an iterative procedure

An equally important contribution of our paper is to address the sensitivity of the cross-sectional fit of no-arbitrage models to measurement errors. We develop a cross-sectional estimator that locally averages out the measurement errors around the date of the model fit. Specifically, we employ a rolling sample approach for estimating the preference parameters from a set of adjacent cross-sections of bond prices. This procedure can be interpreted as a hybrid of cross-sectionally calibrating a no-arbitrage model and estimating the same model from a historical panel of bond prices. Depending on the degree of measurement error in the data, which we specify with a tuning parameter, our estimator resembles more the cross-sectional calibration or panel estimation approach.

To illustrate some of the key features of our approach, we present an empirical application of two-factor models to U.S. Treasury data. We use as factors either consumption growth and inflation, which arise naturally as state variables in any consumption-based asset pricing model, or the level and conditional volatility of the one-month interest rates, which resembles the Longstaff and Schwartz (1992) model. In both cases, the results illustrate clearly the trade-off between fitting as well as possible and overfitting the cross-section of bond prices due to measurement errors. After optimizing this trade-off, by using a quadratic log pricing kernel and a three-month rolling-sample window to average out measurement errors, our approach fits almost exactly the cross-section of bond prices at each date (with mean absolute errors ranging from five to 9 basis points) and produces out-of-sample forecast errors that beat a random walk benchmark and are comparable to the results in the affine term structure literature (with one-year ahead mean absolute forecast errors of 70 to 80 basis points). The performance of our approach is particularly encouraging for the macroeconomic factors, which are known to be difficult to relate to asset prices. Finally, the fact that we find non-linearities in the pricing kernel to be important lends support to quadratic term structure models (Ahn, Dittmar, and Gallant, 2001).

We describe our theoretical framework in Section 2 and then discuss our econometric approach in Section 3. Section 4 presents and interprets our empirical application. We conclude in Section 5 with a review of our approach and empirical results.

2 Theoretical Framework

2.1 A Generic Discrete-Time Term Structure Model

We consider a discrete-time economy with complete markets. The state of the economy is described by $K_z < \infty$ exogenous factors z_t governing endowments and $K_\theta < \infty$ parameters θ_t describing aggregate preferences that can be time-varying (e.g., habits formation). Following Harrison and Kreps (1979), there exists a nominal pricing kernel $M(z_t, z_{t+1}; \theta_t)$ that prices all assets, in the sense that:

$$P_t^x = \mathrm{E}_t\big[M(z_t, z_{t+1}; \theta_t)X_{t+1}\big], \tag{1}$$

where X_{t+1} is the uncertain payoff of the asset at time $t+1$, P_t^x is its price at time t, $\mathrm{E}_t[\cdot]$ denotes an expectation conditional on all information at time t. Furthermore, the absence of arbitrage in this economy is equivalent to the existence of a strictly positive pricing kernel, such that $M(z_t, z_{t+1}; \theta_t) > 0$ for all z_{t+1}.

In a term structure setting, the price P_t^n of a nominal default-free discount bond that pays one dollar when it matures at time $t+n$ is:

$$
\begin{aligned}
P_t^n &= \mathrm{E}_t\big[M(z_t, z_{t+1}; \theta_t)P_{t+1}^{n-1}\big] \\
P_t^n &= \mathrm{E}_t\Big[M(z_t, z_{t+1}; \theta_t)\,\mathrm{E}_{t+1}\big[M(z_{t+1}, z_{t+2}; \theta_{t+1})P_{t+2}^{n-2}\big]\Big] \\
&\quad \cdots \\
&= \mathrm{E}_t\Big[\prod_{j=0}^{n-1} M(z_{t+j}, z_{t+j+1}; \theta_{t+j})\Big],
\end{aligned}
\tag{2}
$$

where the last equality follows by recursive substitution and the law of iterated expectations.

Equation (2) shows that any term structure model in this general setting amounts to parametric assumptions about the pricing kernel $M(z_t, z_{t+1}; \theta_t)$ and the *joint* dynamics of the exogenous factors z_t and preference parameters θ_t. We assume that z_t and θ_t evolve as a first-order Markov process with time-homogeneous transition density:[2]

$$f(z_{t+1}, \theta_{t+1}|z_t, \theta_t; \varphi) = f_z(z_{t+1}|z_t; \varphi_z) \times f_\theta(\theta_{t+1}|z_t, \theta_t; \varphi_\theta), \tag{3}$$

where the factorization of the density follows from the fact that the factors z_t are exogenous. Notice that if the parameters of the pricing kernel are constant (i.e., $\theta_t = \theta$ for all t), as they

---

[2]The first-order assumption is innocuous because z_t and θ_t can contain lagged values.

are in most models, the joint dynamics reduce to $f_z(z_{t+1}|z_t; \varphi_z)$, a first-order Markov process for the exogenous factors alone. We refer to φ as the "time-series parameters" because they characterize the time-series dynamics of the factors and preference parameters.

2.2 Time-Consistency

By time-consistency we mean that the expectation in equation (2) is taken with respect to the distribution governing the actual joint evolution of $\{z_t, \theta_t\}$ in the data (looking forward).[3] Given the factorization in equation (3), time-consistency requires that the transition density $f_z(z_{t+1}|z_t; \varphi_z)$ describes the true dynamics of the factors z_t and that the transition density $f_\theta(\theta_{t+1}|z_t, \theta_t; \varphi_\theta)$ accurately reflects the way the preference parameters θ_t change from one date to the next. Since the factors are assumed observable, the first part of this condition is straightforward to implement and test using historical data $\{z_t\}_{t=1}^T$. An important feature of our econometric procedure is to ensure that the second part also holds.

The problem of time-inconsistency arises in the context of no-arbitrage models. As with most term structure models, the bond pricing equations of no-arbitrage models assume time-invariant factor dynamics and constant preference parameters, meaning that the expectation in equation (2) is taken only with respect to the factor dynamics. The goal of no-arbitrage models is to fit *exactly* the cross-section of bond prices $\{P_t^n\}_{n=1}^N$ at a given date. This perfect cross-sectional fit is typically achieved through the choice of both the preference and time-series parameters. Given enough parameters, the resulting model precludes arbitrage in the cross-section of bonds, but the model is potentially time-inconsistent for two reasons. First, there is no guarantee that the factor dynamics implied by the cross-sectional fit correspond to the true dynamics of the factors, since the model is never confronted with historical data. Second, if the fitted parameters are not constant from one period to the next, the fit is inconsistent with the assumption of constant parameters underlying the bond pricing equations. This implies that the model is internally inconsistent (Dybvig, 1989) and may permit arbitrage in contingent claims (Backus, Foresi, and Zin, 1998).

---

[3]An alternative way to define time-consistency is under the risk neutral measure, defined by the equivalent representation of the pricing equation $P_t^x = E_t[M(z_t, z_{t+1}; \theta) X_{t+1}] = E_t^Q[X_{t+1}]$, where the expectation $E_t^Q[\cdot]$ is taken with respect to the marginal utility discounted density $M(z_{t+1}, z_t; \theta_t) \hat{f}(z_{t+1}, \theta_{t+1}|z_t, \theta_t; \varphi)$. Since the bond pricing equation involves only the *product* of $M(\cdot)$ and $f(\cdot)$, any model with *arbitrary* pair $\hat{M}(z_{t+1}, z_t; \theta_t)$ and $\hat{f}(z_{t+1}, \theta_{t+1}|z_t, \theta_t; \varphi)$ that satisfy $\hat{M}(\cdot)\hat{f}(\cdot) = M(\cdot)f(\cdot)$ in all states is time-consistent under the risk neutral measure. Clearly any model that is time-consistent under the real measure is also time-consistent under the risk neutral measure, but the converse is generally not true.

3 Econometric Approach

The basic idea of our econometric approach is to hold constant the time-series parameters across all cross-sectional fits, enforcing the assumption of time-invariant factor dynamics, and to explicitly incorporate the way the preference parameters change from one fit to the next into the bond pricing equations, effectively taking the expectation in equation (2) with respect to $f(z_{t+1}, \theta_{t+1}|z_t, \theta_t; \varphi)$ as opposed to just $f_z(z_{t+1}|z_t; \varphi_z)$. Specifically, we first estimate the factor dynamics from the historical data $\{z_t\}_{t=1}^T$. We then iterate between inferring the preference parameters θ_t from the cross-section of bond prices at each date t in our sample and estimating the dynamics $f_\theta(\theta_{t+1}|z_t, \theta_t; \varphi)$ from the resulting time-series of inferred preference parameters $\{\theta_t\}_{t=1}^T$. We iterate between these steps because we need the dynamics of the preference parameters to obtain theoretical bond prices from equation (2) for the cross-sectional fits and we need the time-series of preference parameters to estimate their dynamics. We now turn to an example to further illustrate our approach.

3.1 Illustrative Example

Consider the classic Vasicek (1977) model, which in our setting implies the following linear log pricing kernel and Gaussian factor dynamics (Backus, Foresi, and Telmer, 1999):[4,5]

$$m(z_t, z_{t+1}; \theta) = \theta_0 + \theta_1 z_t + \theta_2 z_{t+1} \tag{4}$$

and

$$f_z(z_{t+1}|z_t; \varphi_z) = \mathrm{N}\big[\pi_{z0} + \pi_{z1} z_t, \Sigma_z\big]. \tag{5}$$

With constant preference parameters, the pricing kernel is log-normally distributed and the conditional expectation in the bond pricing equation (2) can be solved analytically in the form $P_t^n = \exp\{A_n(\theta, \varphi_z) + B_n(\theta, \varphi_z) z_t\}$.

The traditional way of calibrating equations (4) and (5) as a no-arbitrage model involves choosing at date t both the time-series and preference parameters to minimize, for example, the sum of squared pricing errors $\sum_{n=1}^N \big[P_t^n - \exp\{A_n(\theta_t, \varphi_{zt}) + B_n(\theta_t, \varphi_{zt}) z_t\}\big]^2$, where here the time-subscripts on the parameters denote the date of the cross-sectional fit. The model is arbitrage-free if this cross-sectional fit is perfect. However, unless $f(z_t|z_{t-1}; \varphi_{zt})$ happens to describes the true dynamics of the factors into the future and $\{\varphi_{zt}, \theta_t\}$ are the same across

---

[4]We work with the log pricing kernel $m(z_t, z_{t+1}; \theta) \equiv \ln M(z_t, z_{t+1}; \theta)$ to ensure positivity of the pricing kernel, which in turn guarantees the absence of admit arbitrage (Harrison and Kreps, 1979).

[5]In the single-factor Vasicek model, the factor z_t is the one-period yield $-\ln P_t^1 = -\ln \mathrm{E}_t[M(z_t, z_{t+1}; \theta)]$. As we discuss in Section 3.3.3, choosing a bond yield as factor is problematic in our setting.

all cross-sectional fits, the model is time-inconsistent.

In the context of this example, our econometric approach proceeds as follows. We first estimate the time-series parameters φ_z corresponding to the dynamics of the exogenous factors from historical data $\{z_t\}_{t=1}^T$. We then choose for each date t in the sample the preference parameters to minimize $\sum_{n=1}^N \left[P_t^n - P^n(z_t, \theta_t, \varphi)\right]^2$, where $P^n(z_t, \theta_t, \varphi)$ denotes the theoretical bond price obtained from equation (2). Since the preference parameters vary from one date to the next, we model their dynamics analogous to equation (5) as:

$$f_\theta(\theta_{t+1}|\theta_t, z_t; \varphi_\theta) = \mathrm{N}\left[\pi_{\theta 0} + \pi_{\theta 1} z_t + \pi_{\theta 2} \theta_t, \Sigma_\theta\right] \tag{6}$$

and evaluate the expectation in equation (2) with respect to the *joint* distribution of the factors and preference parameters, given by the product of equations (5) and (6). The challenge is how to estimate the time-series parameters φ_θ corresponding to the dynamics of the preference parameters, since the preference parameters are only revealed through the cross-sectional fits, which, in turn, depend on the theoretical bond prices and hence on the time-series parameters we want to estimate. We propose an iterative procedure. Starting by taking expectations in the bond pricing equation only with respect to the factor dynamics (as if the preference parameters were constant), we infer the preference parameters for each date in the sample. Given the resulting $\{\theta_t\}_{t=1}^T$, we estimate φ_θ. We then return to inferring the preference parameters from the cross-sections of bond prices, this time taking the expectation in the bond pricing equation with respect to the joint distribution of the factor and preference parameters. We iterate until the distribution of the preference parameters has converged.

While this example illustrates the basic steps of our econometric approach, our actual implementation differs in a number of important ways. First, we use more flexible statistical models for the dynamics of the factors and preference parameters because there is no guarantee that the true dynamics are Gaussian and because any misspecification of the dynamics may lead to time-inconsistencies. Second, we use a more flexible parameterization of the log pricing kernel that provides enough degrees of freedom to fit as well as desired (which means *exactly* in an ideal world without data imperfections) the cross-section of bond prices at each date in our sample. Third, we evaluate the theoretical bond prices using simulations (i.e., by Monte Carlo integration) because with flexibly parameterized dynamics and log pricing kernel the bond pricing equations do not have analytical solutions. Finally, but perhaps most innovative from an econometric perspective, we make the cross-sectional inference of the preference parameters at each date robust to measurement errors in bond prices. We now describe in more detail each step of our econometric procedure.

3.2 Details of the Procedure

3.2.1 Step 1: Estimating the Factor Dynamics

Since by assumption the factors are observable and their dynamics do not depend on the preference parameters or bond prices, we can directly estimate $f_z(z_{t+1}|z_t; \varphi_z)$ from historical data $\{z_t\}_{t=1}^T$. To avoid misspecification, which may lead to time-inconsistencies, we rely on the semi-nonparametric (SNP) density estimation method of Gallant and Tauchen (1989). Intuitively, we model the dynamics of the factors as a nonlinear and potentially time-varying transformation of a Gaussian vector autoregression (VAR). Gallant and Nychka (1987) show that under reasonable regularity conditions, SNP transition densities are flexible enough to approximate arbitrarily well the true transition densities of a stationary Markov process. In addition, the convergence of the SNP densities to the true densities is sufficiently strong to consistently evaluate expectations of functions of z_{t+k}, such as in equation (2), by integrating against the estimated SNP densities instead of against the true densities.

A generic representation of the SNP transition density in our context is:

$$f_z(z_{t+1}|z_t; \varphi_z) \propto \Big(\mathcal{H}\big[u_{t+1}, z_t; \phi_z\big]\Big)^2 \mathrm{N}\big[\pi_{z0} + \pi_{z1}z_t, \Sigma_z\big], \tag{7}$$

where π_{z0} and π_{z1} are the intercept and slope coefficients of a first-order VAR of z_t and Σ_z is the residual covariance matrix of this VAR. $\mathcal{H}[u_{t+1}, z_t, ; \phi_z]$ denotes a Hermite polynomial expansion in the standardized residuals $u_{t+1} = \Sigma_z^{-1/2}(z_{t+1} - \pi_{z0} - \pi_{z1}z_t)$ with coefficients that are themselves polynomials of z_t, to accommodate time-variation in the departure of the true factor dynamics from the VAR (including conditional heteroscedasticity), and ϕ_z are the parameters of the second set of polynomials (in z_t).[6,7] Finally, φ_z now represents the collection of coefficients of the VAR and the polynomial expansion $[\pi_{z0}, \pi_{z1}, \Sigma_z, \phi_z]$.[8] Estimation of the SNP density entails a standard maximum likelihood procedure together with a model selection strategy, such as minimizing the Akaike (1973) information criterion

---

[6] For example, assume there is only one factor and consider a second-order expansion in the standardized residual u_{t+1} with coefficients that are linear in z_t:

$$\mathcal{H}[u_{t+1}, z_t; \phi_z] = a_t + b_t(2u_{t+1}) + c_t(4u_{t+1}^2 - 2),$$

where the terms in parentheses are first- and second-order members of the Hermite class and $a_t = a_0 + a_1(2z_t)$, $b_t = b_0 + b_1(2z_t)$, and $c_t = c_0 + c_1(2z_t)$. In this case, $\phi_z = [a_0, a_1, b_0, b_1, c_0, c_1]$.

[7] One of the regularity conditions for convergence is that the tails of the transition densities of the standardized residuals u_{t+1} are thin enough for $\exp\{u_{t+1}^2/2\}f'(u_{t+1}|\hat{z}_t; \varphi_z)^2$ to be integrable (Cramér, 1925). An intuitive way to ensure convergence for highly heteroskedastic processes, which can have extremely fat tails, is to standardize the data by a *conditional* covariance matrix Σ_{zt} (see Gallant and Tauchen, 1992).

[8] Gallant and Nychka (1987) and Gallant and Tauchen (1989) provide a more detailed description of the SNP density estimator and its asymptotic properties (as the order of the expansion increases). The idea of approximating a density through a Hermite polynomial expansion dates back at least to Schwartz (1967).

(AIC), to determine the appropriate order of the polynomial expansion.

The flexibility and convergence properties of the SNP densities are obviously appealing from a statistical perspective. Moreover, the structure of the SNP densities is ideal for our application from an economic perspective because it nests at the most fundamental level the factor dynamics of the classic Vasicek model in equation (5) (through the VAR). The polynomial expansion then capture both unconditional and conditional deviations of the true factor dynamics from this intuitive and standard (in the literature) reference model.

3.2.2 Step 2: Inferring the Preference Parameters

Fitting arbitrarily well the cross-section of bond prices at each date is likely to require considerable degrees of freedom. Since we hold the time-series parameters constant to enforce the assumption of time-homogenous dynamics, the only way to introduce these degrees of freedom is through the pricing kernel. We therefore model the log pricing kernel also semi-nonparametrically using a Hermite polynomial expansion in z_t and z_{t+1} of the generic form $m(z_t, z_{t+1}; \theta_t) = \mathcal{H}(z_t, z_{t+1}; \theta_t)$, where θ_t represents the coefficients of the expansion.[9] Under suitable regularity conditions for the true pricing kernel, this polynomial expansion converges to the true pricing kernel as the order of the expansion increases (Powell, 1981).

A generic expansion of the log pricing kernel achieves the goal of providing enough degrees of freedom, but it does so along some dimensions that are not particularly sensible for pricing kernels. The pricing kernels generated by von Neumann-Morgenstern preferences can always be expressed in time-separable form $m(z_t, z_{t+1}; \theta_t) = \theta_{0t} + m_1(z_t; \theta_{1t}) + m_2(z_{t+1}; \theta_{2t})$, which eliminates all interactions between the current and future realizations of the factors and implies that the corresponding cross-terms in the polynomial expansion are nuisance terms.[10] In particular, all reduced-form term structure modes in the literature that we are aware of, including the popular classes of affine (Duffie and Kan, 1996) and quadratic (Ahn, Dittmar, and Gallant, 2001) models, are time-separable. We incorporate this theoretical restriction of time-separability by considering only separable expansions of the form:

$$m(z_t, z_{t+1}; \theta_t) = \theta_{0t} + \mathcal{H}(z_t; \theta_{1t}) + \mathcal{H}(z_{t+1}; \theta_{2t}). \tag{8}$$

---

[9]Similar expansions of the pricing kernel are considered in different contexts by Chapman (1997) and Dittmar (2002). An alternative is the neural network approach of Bansal and Viswanathan (1993).

[10]This time-separability property follows directly from the Euler equations defining the pricing kernel and holds even for time non-separable preferences, such as habit formation (e.g., Constantinides, 1990), after suitably redefining the state space. The Euler equations reveal further that the pricing kernel can always be expressed in symmetric form with $m_1(z; \theta_1) = -m_2(z; \theta_2)$ and $\theta_1 = \theta_2$. However, this symmetry property is very dependent on the correct representation of the state space (e.g., it breaks down when we add a constant to z_t). We thank George Constantinides for these insights on the properties of pricing kernels.

Finally, notice that, analogous to the SNP factor dynamics, the expansion the log pricing kernel nests as special case the linear pricing kernel of the Vasicek model in equation (4). The higher-order terms in the expansion reflect deviations of the true pricing kernel from the pricing kernel of this popular reference model.

Given this specification of the log pricing kernel, we infer at each date t in the sample the preference parameters θ_t from the cross-section of bond prices $\{P_t^n\}_{n=1}^N$. For pedagogical reasons we first briefly discuss this cross-sectional inference step assuming that there are no measurement errors in bond prices. Of course, it is quite naive to think that the bond data is perfect, especially since zero-coupon bond prices are usually *estimated* from a much larger cross-section of coupon-bearing bond prices (e.g., McCulloch, 1971).[11] Since ignoring measurement errors is likely to result in over-fitting of the pricing kernel to the cross-section of bond prices, which may lead to time-inconsistencies, the main discussion revolves around the more realistic case involving measurement errors.

Without Measurement Errors

We first assume that the bond prices used to calibrate the model are measured without errors. For a given dynamics of the state variables $f_z(z_{t+1}|z_t; \varphi_z)$ and of the preference parameters $f_\theta(\theta_{t+1}|z_t, \theta_t; \varphi_z)$ (which we discuss in the next section) we infer the parameters of the log pricing kernel from the cross-section of bond prices at each date in the sample. Specifically, define the pricing errors at date t as:

$$
g(z_t, \theta_t) = \begin{bmatrix} P_t^1 - \mathrm{E}_t\left[\exp\left\{ m(z_t, z_{t+1}; \theta_t) \right\} \right] \\ P_t^2 - \mathrm{E}_t\left[\exp\left\{ m(z_t, z_{t+1}; \theta_t) + m(z_{t+1}, z_{t+2}; \theta_{t+1}) \right\} \right] \\ \ldots \\ P_t^N - \mathrm{E}_t\left[\exp\left\{ \sum_{j=0}^{N-1} m(z_{t+j}, z_{t+j+1}; \theta_{t+j+1}) \right\} \right] \end{bmatrix}. \tag{9}
$$

As long as the number of bonds in the cross-section is sufficiently large (e.g., equals the number of parameters in the expansion of the log pricing kernel), we can infer the preference parameters θ_t simply by inverting $g(z_t, \theta_t)$.

In principle, the conditional expectations in equation (9) are taken with respect to the *true n-step* ahead transition densities. However, in practice we evaluate the expectations with respect to estimated analogs $f_z(z_{t+n}|z_t; \hat{\varphi}_z)$ and $f_\theta(\theta_{t+n}|z_t; \hat{\varphi}_\theta)$. Since with the SNP

---

[11]An obvious solution to this problem is to use the coupon-bearing bonds in the estimation. Unfortunately, this is computationally formidable because of the large number of outstanding bonds. Furthermore, bid-ask spreads and infrequent trading cause measurement errors even in coupon-bearing bond prices.

estimators the n-step ahead transition densities do not possess closed-form representations for $n > 1$, we cannot evaluate the expectations analytically and instead evaluate them numerically through conditional (on z_t and θ_t) simulations. In particular, conditional on the observed z_t and trial values for θ_t, we simulate S hypothetical paths of the factors and preference parameters $\{z_{t+1}^s, z_{t+2}^s, ...z_{t+N}^s\}_{s=1}^S$ and $\{\theta_{t+1}^s, \theta_{t+2}^s, ...\theta_{t+N}^s\}_{s=1}^S$ according to the estimated one-step ahead transition densities and then evaluate the conditional expectations with the corresponding sample averages over the simulated paths.[12] We thereby construct the following simulation analog of the conditional moments (2):

$$
g_S(z_t, \theta_t) = \begin{bmatrix}
P_t^1 - \frac{1}{S} \sum_{s=1}^S \exp \left\{ m(z_t^s, z_{t+1}^s; \theta_t) \right\} \\
P_t^2 - \frac{1}{S} \sum_{s=1}^S \exp \left\{ m(z_t^s, z_{t+1}^s; \theta_t) + m(z_{t+1}^s, z_{t+2}^s; \theta_{t+1}^s) \right\} \\
\cdots \\
P_t^N - \frac{1}{S} \sum_{s=1}^S \exp \left\{ \sum_{j=0}^{N-1} m(z_{t+j}^s, z_{t+j+1}^s; \theta_{t+j}^s) \right\}
\end{bmatrix},
\tag{10}
$$

where z_{t+j}^s and θ_{t+j}^s denote the realization of the factors and preference parameters at time $t+j$ along the sth simulated path from the estimated transition densities.

We define the cross-sectional estimator $\hat{\theta}_t$ of the preference parameters θ_t as the parameter vector that sets $g_S(z_t, \theta_t)$ to zero. Repeating this cross-sectional inference step for every date t in the sample generates the sequence of preference parameters $\{\hat{\theta}_t\}_{t=1}^T$ and the corresponding sequence of log pricing kernels $\{m(z_t, z_{t+1}; \hat{\theta}_t)\}_{t=1}^T$. The asymptotic properties of the estimator are relatively straightforward. The key is to recognize that the moment conditions are exact, in contrast, for example, to the noisy *sample* moments of a GMM estimator. There is no sampling variation in the conditional moments (9) and there is only simulation-induced variation in the simulated conditional moments (10). As $S \to \infty$, the simulation errors become negligible and the only randomness in the cross-sectional estimator comes from the estimation of the transition densities. Consequently, the standard errors of $\hat{\theta}_t$ can be obtained by the *delta method* from the standard errors of $\hat{\varphi}$.

With Measurement Errors

We now consider the more realistic case involving measurement errors in bond prices (which, for obvious reasons, is also the focus of our empirical application). Specifically, we consider

---

[12]See Gallant and Tauchen (1992) for a description of the simulation algorithm for SNP densities.

contaminated bond prices of the form:[13]

$$P_t^n = P_t^{n*} + \varepsilon_t^n, \quad \text{for} \ \ n = 1, 2, \ldots N, \tag{11}$$

where P_t^{n*} denotes the true but unobservable bond prices for which equation (2) holds and ε_t^n are zero-mean measurement errors that are iid through time but potentially correlated across bonds. P_t^n are the observed bond prices. The immediate implication of measurement errors in observed bond prices is that the moment conditions (9) hold only in expectation:

$$E_t^\varepsilon[g(z_t, \theta_t)] = 0, \tag{12}$$

where the expectation is taken with respect to the distribution of the errors ε_t.

If the preference parameters are constant (i.e., $\theta_t = \theta$ for all t) we can estimate them through unconditional moments obtained by taking unconditional expectations of the moments (9) and applying the law of iterated expectations. The unconditional expectation effectively averages out the measurement errors through time. Unfortunately, this approach does not work if the preference parameters are time-varying and, more importantly, the unconditional model looses its interpretation of being arbitrage-free unless we are willing to attribute *all* of the mispricing to measurement errors, meaning $\varepsilon_t \equiv g(z_t, \theta)$.

To overcome the problem of measurement errors without sacrificing the no-arbitrage aspect of our model, we adopt a rolling sample approach very similar to that used by Foster and Nelson (1996) for estimating conditional variances. In particular, we define the following rolling sample estimator of the conditional expectation in equation (12):

$$\bar{g}^\tau(z, \theta_t) = \sum_{j=-\tau}^{\tau} \omega(j) \, g(z_{t+j\Delta t}, \theta_t), \tag{13}$$

where $\omega(j)$ denotes a sequence of non-negative weights that are non-increasing in $|j|$. To understand this estimator, suppose $\omega(j) = 1/(2\tau)$ and $\Delta t = 1$. Then, $\bar{g}^\tau(z, \theta_t)$ is a rolling sample average of the pricing errors $g(z_{t+j}, \theta_t)$, where the sample ranges from $t - \tau$ to $t + \tau$. For this average to converge to the conditional expectation of the pricing errors, instead of to the unconditional expectation, as $\tau \to \infty$, we introduce the sampling frequency Δt. Following the arguments of Foster and Nelson (1996), if we increase both this sampling frequency ($\Delta t \to 0$) and the number of pricing errors we average over ($\tau \to \infty$) in such a way that the sample window becomes increasingly focused on date t ($2\tau\Delta t \to 0$), the rolling

---

[13]Alternatively, we can assume multiplicative measurement errors, such that $P_t^n = P_t^{n*}\varepsilon_t^n$, and express the moment conditions (9) in terms of yields or log prices $g(z_t, \theta_t) = \{\ln P_t^n - \ln E_t[M(z_t, z_{t+n}; \theta_t)]\}_{n=1}^N$.

sample average $\bar{g}^\tau(z, \theta_t)$ converges to the conditional expectation $\mathrm{E}_t^\varepsilon[g(z_t, \theta_t)]$.[14]

Intuitively, our estimator averages out the measurement errors over cross-sections of pricing errors surrounding date t, instead of over all cross-sections in the sample (which corresponds to unconditional estimates of constant preference parameters). The purpose of the weights $\omega(j)$ is to emphasize pricing errors close to (in time) date t relative to those far from date t. There are many candidate weighting functions, including an optimal one, that accomplish this.[15] For the sake of simplicity, we focus on the following two specifications

$$\omega(j) = 1 - \frac{|j|}{\tau + 1} \quad \text{or} \quad \omega(j) = \begin{cases} 1 - \dfrac{|j|}{\tau + 1} & \text{if } j \leq 0 \\ 0 & \text{otherwise} \end{cases}. \tag{14}$$

The first set of weights defines a symmetric "tent" centered at $j = 0$ (or date t). It generates a sample window that is both forward- and backward-looking. The second set of weights is equal to the first for $j \leq 0$ but is zero for $j > 0$. It implies a backward-looking sample window that is particularly useful for out-of-sample testing of the estimated pricing kernel.

We use the rolling sample average (13) with weights (14) and the simulated moments (10) to construct the cross-sectional GMM estimator:[16]

$$\hat{\theta}_t^\tau = \arg\min_\theta \bar{g}_S^\tau(z, \theta)' W \bar{g}_S^\tau(z, \theta), \tag{15}$$

for some predetermined weighting matrix W. When $\tau = 0$ we have $\bar{g}_S^0(z, \theta) = g_S(z_t, \theta)$ and hence $\hat{\theta}_t^0 = \hat{\theta}_t$, which means that the formulation of our estimator with measurement errors nests the case without measurement errors. At the other extreme, when $\tau = \infty$, we have $\bar{g}_S^\infty(z, \theta) = \sum_{t=1}^T g_S(z_t, \theta)$, which is proportional (by a factor $1/T$) to the sample analog of the unconditional expectation of the moments (9). Therefore, our approach also nests the unconditional estimator of constant preference parameters. Finally, note that the presence of the weighting matrix W allows for there to be more bonds than the dimension of θ. This flexibility is useful in the presence measurement errors because the information contained in the additional bonds is not necessarily redundant.

There are two natural economically motivated weighting matrices W in our context.[17] The

---

[14]In our notation, $\Delta t \to 0$ means that we are sampling the data at an increasing frequency (weekly, daily, hourly, minute-by-minute, etc.). The natural limit is a continuous record of the data, which is why the asymptotics we invoke below are called "continuous record asymptotics."

[15]Foster and Nelson show that for estimating conditional variances and covariances the optimal, in an asymptotic mean-squared error sense, weighting function is $\omega(j) = \alpha \exp\{-\alpha|j|\}$ for some $\alpha \geq 0$.

[16]The estimator can be interpreted in the context of the simulated method of moments (SMM) of Ingram and Lee (1991) and Duffie and Singleton (1993).

[17]There also exists an optimal weighing matrix that minimizes the asymptotic variance of the estimates

first is a diagonal matrix with elements $1/n$, where $n = 1, 2, \ldots N$, along the diagonal. With this weighting matrix our approach approximates least squares estimation in yields, since $\ln P_t^n / n \simeq (P_t^n - 1)/n$ for $P_t \simeq 1$. By treating equally yield errors across maturities this matrix emphasizes more heavily dollar price errors at the short end of the term structure. The second natural weighting matrix is an identity matrix, which gives the interpretation of least squares estimation in prices. This matrix treats equally dollar price errors across maturities and therefore emphasizes more yield errors for long maturities. As in a standard regression context, variation in the regressors, in our case variation in the future realizations of factors and preference parameters, increases the precision of the estimated parameters. Given the autoregressive nature of the dynamics typically used in the term structure literature, the variance of $\{z_{t+1}^s, \theta_{t+1}^s\}_{s=1}^S$ in the moment condition for the one-period bond is substantially less than that of $\{z_{t+N}^s\}_{s=1}^S$ in the corresponding moment condition for the N-period bond. Therefore, we choose the identity matrix to weight the moment conditions.[18]

Relative to the case without measurement errors, the asymptotics of $\hat\theta_t^\tau$ are substantially more complicated because of the rolling sample average. Not only do we need to ensure that the local (in time) average $\bar g_S^\tau(z, \theta_t)$ converges to the conditional expectation (12), but the sampling variation in the rolling sample average due to the measurement errors also introduces additional randomness in the cross-sectional estimator of the preference parameters. Given our additive measurement error specification (11) and under appropriate regularity conditions, the asymptotic distribution of the rolling sample average estimator as $S \to \infty$, $\tau \to \infty$, $\Delta t \to 0$, and $\Delta t \tau \to 0$ is (Foster and Nelson, 1996):

$$\sqrt\tau \, \bar g_S^\tau(z, \theta_t) \sim \mathrm{N}\big[0, V_\varepsilon\big]. \tag{16}$$

where V_ε denotes the covariance matrix of the measurement errors. To derive the asymptotics of $\hat\theta_t^\tau$ we expand the first-order conditions of the GMM problem (15) as follows:

$$\sqrt\tau\big(\hat\theta_t^\tau - \theta_t\big) = -A^{-1}\left[Dg_\theta' W \sqrt\tau \, \bar g_s^\tau(z, \theta_t) + \big[Dg_\theta' W Dg_\varphi\big]\frac{\sqrt\tau}{\sqrt T}\sqrt T(\hat\varphi - \varphi)\right]. \tag{17}$$

Finally, substituting the asymptotic distributions of $\bar g_S^\tau(z, \theta_t)$ and of $\hat\varphi$ into this expansion reveals that as $T \to \infty$ and $\sqrt\tau/\sqrt T \to c$ (a constant) the estimator $\hat\theta_t^\tau$ is asymptotically

---

(Hansen, 1982). However, we prefer sacrificing some statistical efficiency in return for the transparency and intuition offered by an economically motivated weighting matrices (Cochrane, 2001).

   [18]This reasoning supports further the use of an economically motivated weighting matrix instead of the statistically optimal weighting matrix (see also footnote 17). The optimal weighting matrix emphasizes the least variable moments, in our case the pricing errors for short-term bonds, which unfortunately are also the least informative moments for identifying the parameters of the pricing kernel.

normal with mean θ_t and covariance matrix:

$$\text{Asy. Var}\big[\hat{\theta}_t^\tau\big] = \frac{1}{\tau} A^{-1} \big[Dg_\theta' W \, V_g \, W' Dg_\theta + c \, Dg_\theta' W Dg_\varphi \, V_\varphi \, Dg_\varphi' W' Dg_\theta \big] A^{-1}. \qquad (18)$$

This asymptotic covariance matrix illustrates that our cross-sectional estimator of the preference parameters is influenced by both the sampling variation in the estimates of the time-series parameters (the second term in the brackets) and the sampling variation in the moment conditions due to the measurement errors (the first term).

3.2.3 Step 3: Estimating the Preference Parameter Dynamics

Given a sequence of preference parameters $\{\theta_t\}_{t=1}^T$ obtained by applying the above cross-section inference step at each date in the sample, we estimate the dynamics of the preference parameters $f_\theta(\theta_{t+1}|z_t, \theta_t; \varphi_\theta)$ analogous to the way we estimate the dynamics of the factors in the first step of the procedure. Specifically, we model the transition densities of the preference parameters using the SNP densities:[19]

$$f_\theta(\theta_{t+1}|z_t, \theta_t; \varphi_\theta) \propto \Big(\mathcal{H}\big[v_{t+1}, [z_t, \theta_t]; \phi_\theta\big] \Big)^2 \text{N}\big[\pi_{\theta 0} + \pi_{\theta 1} z_t + \pi_{\theta 2} \theta_t, \Sigma_\theta\big], \qquad (19)$$

where $v_t = \Sigma_\theta^{-1/2}(\theta_{t+1} - \pi_{\theta 0} - \pi_{\theta 1} z_t - \pi_{\theta 2} \theta_t)$ and $\varphi_\theta = [\phi_\theta, \pi_{\theta 0}, \pi_{\theta 1}, \pi_{\theta 2}, \Sigma_\theta]$. The only practical difference between estimating the factor dynamics and preference parameter dynamics is that the dimensionality of θ_t tends to be much larger than that of z_t. This limits the complexity of the SNP densities we can realistically consider.

3.2.4 Iterating between Steps 2 and 3

The second and third step of our procedure are intimately linked. On one hand, the second step requires estimates of the preference parameter dynamics to evaluate the theoretical bond prices. On the other hand, the third step is based on the sequence of preference parameters from the second step. We accommodate this link by iterating between the two steps.

We initiate the iteration by assuming in the second step that the preference parameters are constant. That is, in the first iteration we assume at each date t that $\theta_{t+n} = \theta_t$ for all future dates n and evaluate the conditional expectations in equation (9) only with respect to the estimated factor dynamics (which may be time-inconsistent). We use the resulting sequence of preference parameters to estimate their dynamics in the third step. We then

---

[19]Notice that we ignore the potential estimation error in the preference parameters from the cross-sectional inference step. The way to incorporate the estimation error is to formulate the preference parameter dynamics in a state-space form. This significantly complicates the estimation and is beyond the scope of this paper.

return to the cross-sectional inference step, this time (and every time thereafter) evaluating the conditional expectations with respect to the estimated joint dynamics of the factors and preference parameters from the previous iteration. Given a new set of preference parameters, we re-estimate their dynamics and then obtain again a new set of preference parameters. We terminate the iteration when the preference parameters are unchanged from one iteration to the next. Formally, we implement the convergence criterion $\max_{t,\dim(\theta)} |\theta_t^{(i)} - \theta_t^{(i-1)}| < \eta$, where i enumerates the iteration and η is a predetermined critical value.

3.3   Discussion

3.3.1   Factors vs. Time-Varying Preference Parameters

Allowing the parameters of the pricing kernel to be time-varying commingles, to some extent, the notions of factors and preference parameters. To clearly differentiate the two, the factors are to be thought of as exogenous to the pricing environment (consumption growth, inflation, GNP growth, etc.), while the preference parameters reflect the attitude of agents toward these factors. An example that fits well into this structure is the habit formation model of Campbell and Cochrane (1999). In their economy, consumption growth rates are independent of preferences (the factor dynamics do not depend on θ_t) and agents realize shocks to their habit level (or relative risk aversion) from a distribution that depends on the consumption growth rates (the preference parameter dynamics depend on z_t).

In practice, when the true structure of the economy is unknown, the danger exists that an incorrect choice of factors leads to spurious variation in the preference parameters. Consider the extreme case where the economy is driven by a set of factors that is completely orthogonal to the factor we use in the estimation. In that case, the variation in the preference parameters fully reflects the state of the economy, rather than aggregate preferences. Of course, more rigidly parameterized models (i.e., model with constant preference parameters) completely break down in this case. We therefore view the flexibility of our approach to accommodate this form of model misspecification as a strength, rather than a weakness, but acknowledge that interpreting the variation in the preference parameters requires some caution.

Another potentially controversial aspect of our modeling approach is that we assume the factors are observable, while the majority of the term structure literature allows at least some of the factors to be latent. However, this assumption is actually less restrictive in our setting than it appears because we are not imposing any particular parametric form on the pricing kernel. The key is that as long as there exists an invertible mapping between the observed and latent factors, we can price assets just as well with the observed factors as with

15

the latent factors. The only difference is that the functional form of the pricing kernel with the observed factors may be different from that with the latent factors.

3.3.2 Approximate vs. Exact No-Arbitrage Model

As long as there exists a pricing kernel that is consistent with the observed bond prices and this pricing kernel is a smooth function, the convergence theorems for polynomial expansions (Powel, 1981) guarantee that our model can in principle deliver a perfect cross-sectional fit and therefore belongs to the class of no-arbitrage models. However, depending on the order of the polynomial we use in practice, the cross-sectional fit of the model can be less than perfect. On one hand, when there are no measurement errors in bond prices, the presence of any pricing errors implies that the model is not completely arbitrage-free and suggests that we have to increase the order of the polynomial expansion. On the other hand, when there are data imperfections, increasing the order of the expansion may lead to over-fitting.

Data imperfections are a reality. This means that there always exists a trade-off between fitting as well as possible and over-fitting the model to the cross-section of bond prices. As a result, the best specification of the pricing kernel, in terms of optimizing this trade-off, necessarily corresponds to a model that is only approximately arbitrage-free (i.e., the model generates some differences between observed and model prices). This should not be viewed as a drawback of our approach. Rather, the flexibility of our approach to achieve a compromise between fitting and over-fitting is one of its major advantages relative to traditional no-arbitrage models that can only be calibrated exactly.

3.3.3 Yield-Based Factors

Instead of using macroeconomic state variables as factors, the majority of the term structure literature uses yield-based factors. In the context of our approach, yield-based factors pose two challenges. First, bond prices or yields do not generally satisfy the factorization of the joint dynamics of the factors and preference parameters in equation (3) because they depend explicitly on the preference parameters through the bond pricing equation (2). Without the factorization of the transition densities we cannot split the estimation of the factor dynamics (in step 1) from that of the preference parameter dynamics (in step 3). A potential remedy, therefore, is to directly estimate the joint dynamics $f(z_{t+1}, \theta_{t+1}|z_t, \theta_t; \varphi)$ in the iterative scheme. Unfortunately, the resulting increase in the dimensionality of the transition densities makes it more difficult to implement higher-order SNP densities.[20]

---

[20]Recall that the number of terms in an expansion increases exponentially in the number of variables.

Second, yield-based factors introduce additional restrictions on the parameters of the pricing kernel arising from the fact that the model must recover exactly the yields used as factors. Suppose, for example, we take the Vasicek model described in equations (4) and (5) and use as factor the one-period bond yield $z_t = -\ln P_t^1$. Since:

$$-z_t = \ln P_t^1 = \ln E_t[\exp\{\theta_0 + \theta_1 z_t + \theta_2 z_{t+1}\}] = \theta_0 + \theta_1 z_t + \theta_2(\pi_{z0} + \pi_{z1} z_t) + \frac{1}{2}\theta_2^2 \Sigma. \quad (20)$$

the intercept of the pricing kernel θ_0 is:

$$\theta_0 = -(1 + \theta_1 + \theta_2 \pi_{z1})z_t - \theta_2 \pi_{z0} - \frac{1}{2}\theta_2^2 \Sigma. \quad (21)$$

The intercept cannot depend on z_t, which implies the following two parameter restrictions $(1 + \theta_1 + \theta_2 \pi_{z1}) = 0$ and $\theta_0 = -\theta_2 \pi_{z0} - \frac{1}{2}\theta_2^2 \Sigma$. This leads to two equations in three unknowns, leaving only one degree of freedom. In other words, two of the three preference parameters in this example are determined by the fact that the factor is the one-period bond yield.

In general, the specifics of the parameter restrictions depend on the parameterizations of the dynamics and pricing kernel, as well as on the identity of the factors, but the generic lesson from this example is that as long as the factors relate to bond prices (perhaps through some nonlinear mapping), the model must recover the factors in all states of the world or, equivalently, for all realizations of the factors. As long as the bond pricing equations (2) have analytical solutions, we can impose these identification restrictions analytically. Otherwise, we need to impose them numerically. Suppose the factors are $z_t = G(\{P_t^n\}_{n=1}^N)$. In that case, we simply add another set of (potentially infinite-dimensional) moments:

$$z - G\left(\left\{E\left[\prod_{j=0}^{n-1} M(z_{t+j}, z_{t+j+1})\Big| z_t = z\right]\right\}_{n=1}^N\right) = 0 \text{ for all possible realizations } z \text{ of } z_t \quad (22)$$

to the estimation moments (9). Theoretically, there are two (asymptotically equivalent) ways to enforce the "for all possible realizations" condition. The first is to impose the identification moments literally for all z in the state-space of z_t (perhaps through random sampling). The second way is to impose the moments for every observed $z = \{z_t\}_{t=1}^T$ in the time-series.[21]

Despite these challenges, one of the empirical applications we provide below involves a two factor model in which one of the factors is the one-month yield. We proceed with this application because the overwhelming predominance of yield-based factors in the term structure literature suggests that there may be considerable interest in implementing our

---

[21] Asymptotically, the time-series of z_t visits the economically relevant regions of the state-space of z_t.

approach with yield-based factors. The application further illustrates the complications that arise and provides some diagnostics of how severe the endogeneity of the factor dynamics is and of how well we can numerically impose the additional parameter restrictions.

4   Empirical Application

We now turn to an empirical application of our econometric approach. Specifically, we apply our procedure to U.S. Treasury data, taking as factors either real consumption growth and inflation or the level and conditional volatility of the one-period interest rate.

4.1   Data

We use an updated version of the McCulloch and Kwon (1993) U.S. Treasury term structure data. This data is a panel of monthly zero-coupon bond prices for maturities ranging from one month to more than 10 years. The zero-coupon bond prices are inferred through a cubic spline discount function fitted to observed coupon-bearing bond prices from the Center for Research in Security Prices (CRSP).[22] For the years 1959 through 1986 we use the original McCulloch and Kwon data, and for the years 1987 through 1998 we use an updated version of the McCulloch and Kwon data from Bliss (1997) that adds five years to the panel.[23]

The data contains 480 observations in the time-series dimension. In the cross-section, we choose 18 bonds with maturities ranging from one month to ten year. Specifically, we use bonds with $\{1, 3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 48, 60, 72, 84, 96, 108, 120\}$ months to maturity. The first two panels of Table 1 provide summary statistics of the yields and yield spreads, relative to the one-month yield, for selected maturities. The following stylized facts emerge: the average yields rise with maturity, ranging from 5.8 percent for a one-month bill to 7.3 percent for a ten-years bond. However, the rate of increase falls with maturity, which implies on average an upward sloping but concave term structure. All yields are persistent, but long-term yields, with autocorrelations of 0.99, are more persistent than short-term yields, with autocorrelations of 0.97. Both the yields and yield spreads are quite volatile, which implies that there is substantial variation in both the level and shape of the term structure. This implication is confirmed by Figure 1, which plots the bond yields.

---

[22]The details of the fitting procedure are provided in McCulloch (1971,1975).

[23]As Bliss (1997) explains in detail, the procedure used for the update is slightly different from the one used by McCulloch and Kwon (1993) in its treatment of the asymptotic rate and the tax effects.

4.2 Factors and Factor Dynamics

We consider two separate sets each with two factors. We focus on two-factor models because there is by now substantial empirical evidence against single-factor models of the *nominal* term structure (Stambaugh, 1988; Litterman and Scheinkman, 1991).[24] As first set of factors, we use macroeconomic variables, which are a natural choice of factors in our framework because they satisfy the requirement of being exogenous to aggregate preferences.[25] We collect data on per-capita real consumption growth, defined as the growth rate of real per-capita consumption of non-durables and services, and inflation, defined as the growth rate of the implicit price deflator for non-durables and services consumption, from DRI/CITIBASE. Consumption growth and inflation are standard state variables in consumption-based term structure models, where the real pricing kernel is a function of aggregate consumption and where inflation is needed to price nominal bonds (Boudoukh, 1993; Brandt and Wang, 2001). Panel C of Table 1 presents summary statistics of these macroeconomic factors.

In order to relate our results to the majority of the term structure literature that uses yield-based factors, we also consider as factors the level and conditional volatility of the one-month interest rate, despite the complications this introduces. The level and conditional volatility of the short rate emerge as state variables in the Longstaff and Schwartz (1992) model, for example, which is a leading example of the popular class of affine term structure models (Duffie and Kan, 1996; Dai and Singleton, 2000). Since the conditional volatility is unobservable, we extract it by augmenting the SNP dynamics with a standard GARCH(1,1) model for the volatility. Panel D of Table 1 describe the yield-based factors.

We estimate SNP dynamics $f_z(z_{t+1}|z_t; \varphi_z)$ for both sets of factors. Based on the AIC, the consumption and inflation factors are best described by VAR with skewed and fat-tailed innovations (a fourth-order expansion of the standardized residuals with constant coefficients). The one-month rate follows a heteroskedatic AR process with innovations that exhibit skewness and excess kurtosis dependent on the level of the rate (a fourth-order expansion with linear coefficients). The point estimates of φ_z are available on request.

---

[24]Litterman and Scheinkman (1991) find that two factors account for 96 percent of the time-series variation in the panel of nominal zero-coupon bond yields. A third factor explains only an additional two percent. Consistent with this finding, Gibbons and Ramaswamy (1993) and Brown and Schaefer (1994) argue that single-factor models can fit the term structure of real instead of nominal interest rates. Finally, Cochrane and Piazzesi (2002) show that expected excess bond returns over a one-year horizon are proportional across bonds with different maturities, suggesting that only one of the factors in bond returns is priced.

[25]Ang and Piazessi (2002) document the importance of macro variables in modeling the term structure.

4.3 Consumption and Inflation Factors

Using first the consumption growth and inflation factors, we apply our iterative procedure, switching between inferring the preference parameters θ_t from the cross-section of bond prices at each date in the sample (with an identity matrix for W and $S = 5000$ simulations) and estimating the preference parameters dynamics $f_\theta(\theta_{t+1}|z_t, \theta_t; \varphi_\theta)$ from the inferred sequence $\{\theta_t\}_{t=1}^T$.[26] After the initial iteration, we model the preference parameter dynamics using a heteroskedatic Gaussian VAR with the factors as exogenous predictors.[27] Using a critical value of $\eta = 0.001$, the procedure converges after four iterations. The point estimates of φ_z from the final iteration are available on request.

4.3.1 In-Sample-Fit

Panel A of Table 2 describes the in-sample fit of linear, quadratic, and cubic log pricing kernels for a pure cross-sectional fit with $\tau = 0$, an unconditional fit with $\tau = \infty$, and two intermediate cases with $\tau = 2$ and $\tau = 5$, which correspond to a three- and six-month rolling sample window to locally average out the measurement errors, respectively. Columns 2–5 of the panel report the mean yield errors, in terms of percent, for bonds at date t using the preference parameters fitted to the cross-sections of bond prices at dates $t - \tau$ to t. In columns 6–9 we present the corresponding mean absolute yield errors. For the top half of the panel we use as factor dynamics a Gaussian VAR and for the bottom half we use the SNP densities chosen according to the AIC (described above).

Consider first the pure cross-sectional fit with $\tau = 0$, which is most comparable to the traditional no-arbitrage approach. Two patterns emerge from the mean and mean absolute yield errors. First, the bias and variability of the in-sample errors decrease for both factor dynamics as the order of the pricing kernel increases. Furthermore, the improvement in fit in going from a first- to second-order pricing kernel is approximately the same as in going from a second- to third-order pricing kernel. For example, with SNP dynamics the mean absolute yield errors of first-, second-, and third-order log pricing kernels are 10.6, 8.8, and 6.5 basis points, respectively. Second, comparing the results for the VAR factor dynamics to those for the SNP dynamics suggests that the factor dynamics are of little, if any, importance for the cross-sectional fit. The difference in mean absolute errors between any two models with the same pricing kernel but different factor dynamics is less than one basis point.

---

[26]We use variance reduction techniques to limit the effect of the simulation errors. In particular, we employ the method of antithetic variates, where for every path simulated from the SNP dynamics with innovations $\{\varepsilon_{t+n}^s\}_{n=1}^N$ we also include the path generated by the innovations $\{-\varepsilon_{t+n}^s\}_{n=1}^N$.

[27]Unfortunately, higher-order specifications with non-Gausian innovations are practically infeasible due to the large number of preference parameters (ranging from five for linear to 19 for cubic pricing kernels).

These patterns are not surprising. In particular, the first pattern is simply a consequence of increasing the number of free parameters. It is quite intuitive and in fact reassuring that increasing the number of free parameters, while holding constant the number of moment conditions, yields an improvement in the fit of the model. The second pattern relates to the issue of time-consistency. It illustrates that the extent to which the factor dynamics are representative of the data does not matter for cross-sectional calibration. This is why standard no-arbitrage models for which both the time-series and preference parameters are calibrated to the cross-section of bond prices can easily imply misspecified factor dynamics.

Comparing the results of the pure cross-sectional fit to those of the unconditional fit with $\tau = \infty$ demonstrates clearly why practitioners prefer calibrating term structure models to the cross-section of bond prices over estimating them from a historical panel. The mean absolute yield error of the best unconditional model is more than six times that of the worst conditional model (76.8 vs. 11.3 basis points). It is almost twelve times as large as the mean absolute error of the best conditional model (76.8 vs. 6.5 basis points). As with the cross-sectional fit, the absolute error decreases with the order of the log pricing kernel. However, the drop is much more dramatic in going from a first- to second-order log pricing kernel than in going from a second- to third-order log pricing kernel. For example, with SNP dynamics the improvement from quadratic terms is 69 basis points but the improvement from cubic terms is only 17 basis points. This reflects a pattern that will become more apparent later, namely, that once over-parameterizing is accounted for, quadratic pricing kernels seem to fit the data best. Finally, the model fits unconditionally much better with SNP than with VAR factor dynamics, which shows that the factor dynamics become important once the model is asked to fit more than a single cross-section of bond prices.

Having digested the results for the two extreme cases, the pure cross-sectional fit and the unconditional fit, understanding the results for the rolling sample estimators with $\tau = 2$ and $\tau = 5$ is straightforward. In both of these intermediate cases, the quality of the in-sample fit improves as we increase the order of the expansion. As in the unconditional case, the improvement in absolute pricing error in going from a first- to second-order log pricing kernel is larger than in going from a second- to third-order log pricing kernel (more so for $\tau = 5$ than for $\tau = 2$). Naturally, the mean absolute pricing errors increase with τ, since the model is asked to price an increasing number of cross-sections of bonds at any point in time. This increase is slightly more dramatic for the models with VAR than with SNP factor dynamics. Furthermore, the difference between the mean absolute pricing errors for the models with VAR and SNP factor dynamics increases slightly as τ increase. Together these two patterns are consistent with the conclusion from the unconditional fit regarding the importance of the factor dynamics for fitting more than a single cross-section of bond prices.

On one hand, someone who views our approach from the perspective of the no-arbitrage literature that ignores the issue of measurement errors may be troubled by the fact that our model produces pricing errors at all, since we advertise our approach as being arbitrage-free. With perfect data, the presence of pricing errors suggests that we need to increase the order of the polynomial expansion of the log pricing error. Indeed, the cross-sectional fit is virtually exact when we consider a fifth-order expansion. On the other hand, someone who buys into our claim that data imperfections are important may be concerned that the statistics we present for the in-sample fit are relatively uninformative about the parameterization of the log pricing kernel and degree of local averaging required to obtain the best compromise between fitting as well as possible and overfitting the data. In order to appease both of these critics, we next examine the ability of each specification to forecast bond prices l periods into the future. With perfect data, the better the cross-sectional fit the better should also be the forecasting performance. In contrast, with less than perfect data, the forecasting results should reveal the parameterization and degree of local averaging that provides the best compromise between fitting and overfitting the data.

4.3.2   Out-of-Sample Forecasts

For each date t in the sample, we use the model fitted to the cross-sections of bond prices at dates $t - \tau$ to t to then price the cross-section of bonds at date $t + l$ (l periods in the future). Panel B in Table 2 reports the resulting mean yield forecast errors (again, in terms of percent). We consider forecast horizons l of three, six, nine, and 12 months.

Consider first the out-of-sample performance of a pure cross-sectional fit with $\tau = 0$. Three clear patterns emerge from the results. First, the mean absolute errors increase with the horizon, which is expected in any forecasting exercise. For example, with a linear log pricing kernel and SNP factor dynamics, the mean absolute errors increase from 50.2 basis points at the three-month horizon to slightly less than one percent at the one-year horizon. Second, the mean absolute errors decrease in the order of the polynomial expansion, approximately five to 10 basis points between the linear and cubic log pricing kernels. This result holds about equally for VAR and SNP factor dynamics and suggests that the superior in-sample fit of the higher-order log pricing kernels is not entirely due to overfitting. Third, the mean absolute errors are significantly smaller with SNP than with VAR factor dynamics, which confirms the importance of correctly specifying the dynamics. For example, at the three-month and one-year horizons the mean absolute errors of a linear model with SNP dynamics are 8.8 and 6.2 basis points less than the corresponding results with VAR dynamics.

The overall magnitude of the forecast errors is large. The mean absolute errors increase

by a factor of five to 10 in going from the in-sample fit to the six-month forecasts. In fact, at the one-year horizon the forecasts of the no-arbitrage models are worse than the fit of the corresponding unconditional models. A significant portion of these large forecast errors, however, are attributed to the sensitivity of the pure cross-sectional fit to measurement errors in bond prices. This can be seen by comparing the results for $\tau = 0$ (a pure cross-sectional fit) to those for $\tau = 2$ (a three-month rolling sample), especially at longer forecast horizons. For instance, at the one-year horizon the mean absolute error of a quadratic model with SNP factor dynamics is 88.9 basis points with $\tau = 0$ and 79.5 basis points with $\tau = 2$.

There is a clear trade-off between fitting the cross-section of bond prices and averaging out the measurement errors, especially at longer horizons. Increasing the size of the rolling-sample from three to six months slightly deteriorates the out-of-sample fit at the nine-month and one-year horizons. Even at shorter horizons, the improvement from increasing the size of the rolling-sample window from three to six months are negligible relative to the improvement from increasing the size from one to three months. Overall, a three-month rolling-sample window appears to achieve a reasonable compromise between fitting the cross-section of bond prices and averaging out the measurement errors.

Focusing on the three-month rolling-sample results with $\tau = 2$, the role of non-linearities in the log pricing kernel is transparent. Introducing quadratic terms in the log pricing kernel results in a significant drop (approximately 5 to 10 basis points) in the mean absolute forecast errors at all horizons and for both factor dynamics. Adding cubic terms, in contrast, yields relatively little improvements (one to two basis points) in the forecasting performance. This result suggests that the true pricing kernel is likely to contain quadratic but not necessarily cubic terms, which is an intriguing finding in light of the recent interest in quadratic term structure models (Ahn, Dittmar, and Gallant, 2001).

The result in Table 2 are aggregated across the whole yield curve. It is not clear, however, that the results that hold in aggregate (the effect of measurement errors on the cross-sectional fit and the role of the non-linearities in the log pricing kernel) apply equally to all bonds in the cross-section. We therefore plot in Figure 2 the mean absolute forecast errors for bonds with one, five, and ten years to maturity as a function of the forecast horizon. The figure is for models with SNP factor dynamics. The circle, diamond, and square symbols are for a first-, second-, and third-order log pricing kernel, respectively. The first row of plots corresponds to a pure cross-sectional fit with $\tau = 0$. The results in the second and third rows are for three- and six-month rolling samples with $\tau = 2$ and $\tau = 5$, respectively. For each case we also plot the mean absolute forecast errors of a simplistic bond pricing model in which yields for a given maturity follow a martingale, so that $\mathrm{E}_t\big[\log(P_{t+1}^n)/n\big] = \log(P_t^n)/n$. This

23

martingale model is an intuitive benchmark. For a term structure model to be valuable out-of-sample, it must at least outperform a forecasting model in which the future yield curve at any horizon is simply equal to the current yield curve.

One prominent feature of the disaggregated results is that the mean absolute forecast errors for the one-year bond are larger than the ones for the five- and ten-year bonds. This observation is not as much a statement about our approach as it is about the data. From Table 1, long-term bond yields are more persistent and less volatile than short-term yields, which makes them easier to forecast. More interestingly, the overall magnitude of the mean absolute forecast errors is comparable to other studies in the literature. For example, our mean absolute forecast errors for the 10-year yield at the three- and six-month horizons of about 38 and 48 basis points are comparable to the forecasting performance of essentially affine term structure models estimated in a more traditional way from a panel of bond prices, reported by Duffee (2002).[28] Although one can produce slightly better forecasts than Duffee using more statistical approaches (Cochrane and Piazzesi, 2002; Diebold and Li, 2002), the fact that our results are comparable to those for essentially affine models is encouraging. This is especially so because we use as factors consumption growth and inflation, which are known to have difficulties relating to asset prices.

Looking at the first row of plots for $\tau = 0$, the martingale benchmark model dominates the linear log pricing kernel at all horizons but is slightly inferior to the quadratic and cubic models at longer horizons. The higher-order expansions perform better than the linear model for all three bonds and at all horizons. However, there is little difference between the results for the quadratic and cubic models. Finally, the differences between the first-, second-, and third-order expansions are slightly more pronounced for long-term bonds than for short-term bonds, which empirically confirms our argument in Section 3.2.2 that long-term bonds are more sensitive to (and hence are more informative about) the pricing kernel.

Comparing the second and third rows to the first clearly illustrates the benefits of locally averaging out the measurement errors in bond prices. The higher-order models significantly outperform the martingale benchmark model for all three bonds at horizons of more than three months. It is also clear from the last two rows that the main improvement in the forecasting performance of the higher-order models comes from including quadric terms in the log pricing kernel. The differences in the results for the quadratic and cubic log pricing kernels are again very small – as the diamond and square symbols are practically on top of each other. This lends further support to our earlier claim that the data seem to favor a

---

[28]Duffee (2002) reports in Table VIII RMSEs of 46 and 62 basis points for three- and six-month forecasts of the 10-year yield. The RMSEs for forecasts of the six-month yield are 28 and 37 and of the two-year yield are 45 and 56 basis points. Our results for the one-year yield are slightly worse.

quadratic pricing kernel. Finally, notice that the forecasting performance of the higher-order models deteriorates for all three bonds when we increase the size of the rolling-sample from $\tau = 2$ in the second row to $\tau = 5$ in the third row. The three-month rolling-sample estimator emerges uniformly as the best compromise between fitting and overfitting the data.

4.4 Yield-Based Factors

Despite the relative success of our approach with macroeconomic factors and the challenges posed by using yield-based factors, we explore an application in which one of the two factors is a bond yield. The goal of this application is to illustrate further the issues that arise with yield-based factors and to present results that are more directly comparable to the literature. As we described in Section 4.2, we use as first factor the one-month yield and as second factor the conditional volatility of the one-month yield extracted by augmenting the SNP dynamics with a standard GARCH(1,1) model.

We implement our approach with these yield-based factors in the same way as with the macroeconomic factors, except that we augment the cross-sectional pricing errors (9) with 10 additional moments of the form (22) for z equal to the deciles of the unconditional distribution of the one-month rate implied by the SNP dynamics. We assign to each of these additional moments a weight that is 10 times larger than the weight assigned to the pricing errors, since the parameter restrictions should hold independent of the cross-sectional fit. However, even with these relatively large weights, the numerical procedure is not exact. The mean absolute difference between the one-month rate deciles and the corresponding model-implied one-month rates ranges from 0.6 to 1.5 basis points, depending on the parameterization of the log pricing kernel. Although this means that, strictly speaking, there is a slight inconsistency between the factors and the model, we proceed with the application because the discrepancy is small by statistical and economic considerations. In fact, it is five to ten times smaller than the in-sample fit with macroeconomic factors.

To measure the severity of the endogeneity problem, which refers to the fact that the yield-based factors may not be exogenous to the preference parameters, we regress changes in the factors from date t to $t+1$ on the factors and preference parameters inferred from the cross-section of bond prices at date t. The preference parameters are individually and jointly insignificant in this regression. Although this does not rule out higher-order dependence between the factors and preference parameters, the evidence supports the factorization of the transition densities in equation (3) for our particular choice of yield-based factors.

25

4.4.1   In-Sample Fit

In Panel A of Table 3 we describe the in-sample fit with yield-based factors and SNP dynamics (the VAR results are very similar and are hence omitted to conserve space). Considering first again the mean absolute errors for the two extreme cases with $\tau = 0$ and $\tau = \infty$, it is clear that as the order of the expansion increases the in-sample fit improves. In particular, the mean absolute errors decline by about two basis points for each additional order of the expansion. Furthermore, comparing the results to those in panel A of Table 2 shows that, for any τ, the yield-based factors produce fitting errors that are substantially smaller than the fitting errors with the macroeconomic factors. For example, in the case of $\tau = 2$, the cubic log pricing kernel leads to mean absolute errors that are eight basis points lower. This is not too surprising as the highly persistent yield-based factors are much more in tune with the way the yield curve behaves than the almost iid macroeconomic factors. As in the case of the macroeconomic factors, the absolute pricing errors increase with τ.

4.4.2   Out-of Sample Forecasts

We present the out-of-sample forecasting results in panel B of Table 3 and in Figure 3. The patterns are very similar to the corresponding results for the macroeconomic factors in panel B of Table 2 and in Figure 2. For horizons of three months and longer, the three-month rolling-sample estimator with $\tau = 2$ generates the best forecasts. In terms of the non-linearities of the log pricing kernel, the majority of the improvement is coming again from the quadratic terms. Including cubic terms in the log pricing kernel yields only a marginal improvement in the forecasting performance. Finally, the yield-based factors produce forecast errors that are substantially lower than the results for the macroeconomic factors. For instance, at the six-month horizon the mean absolute forecast errors are about 10 basis points lower. However, notice that the superiority of the yield-based factors diminishes beyond the six-month forecast horizon. One way to interpret this result is that the macroeconomic factors are picking up low-frequency movements but are not capable of reconciling as well as the yield-based factors the high-frequency fluctuations of the yield curve.

5   Conclusion

We developed an econometric procedure for calibrating no-arbitrage term structure models in a way that is time-consistent and robust to measurement errors. Typical no-arbitrage models are time-inconsistent because their parameters are assumed constant for pricing purposes despite the fact that the parameters change whenever the model is recalibrated and because

the factor dynamics implied by the cross-sectional calibration may not correspond to the actual dynamics of the factors through time. Furthermore, no-arbitrage models are sensitive to measurement errors because they fit exactly each potentially contaminated bond price in the cross-section. We showed how to overcome both problems by evaluating bond prices using the actual joint dynamics of the factors and calibrated parameters and by locally averaging out the measurement errors.

Our econometric procedure achieves time-consistency using an iterative scheme. We first estimate the dynamics of factors governing the endowments in the economy, which are assumed exogenous to the bond market. We then iterate between inferring from the cross-section of bond prices at each data the preference parameters that make the model approximately arbitrage free and estimating the dynamics of the preference parameters from one cross-sectional fit to the next. The key to achieving time-consistency is that the theoretical bond prices used in the first step take into account the dynamics of the preference parameters estimated in the second step. Our procedure explicitly addresses the issue of measurement errors through a rolling sample approach for estimating the preference parameters from the current and adjacent cross-sections of bond prices. This approach can be interpreted as a hybrid of cross-sectional calibration of a no-arbitrage model and estimating the same model from an historical panel of bond prices.

We presented an application to U.S. Treasury data using either macroeconomic or yield-based factors. The results illustrate clearly the trade-off between fitting as well as possible and overfitting the cross-section of bond prices due to measurement errors. After optimizing this trade-off, by using a quadratic log pricing kernel and a three-month rolling-sample window to average out measurement errors, our approach fits almost exactly the cross-section of bond prices at each date (with mean absolute errors ranging from five to 9 basis points) and produces out-of-sample forecast errors that beat a random walk benchmark and are comparable to the results in the affine term structure literature (with one-year ahead mean absolute forecast errors of 70 to 80 basis points). The performance of our approach is particularly encouraging for the macroeconomic factors, which are known to be difficult to relate to asset prices. Finally, the fact that we find non-linearities in the pricing kernel to be important lends support to quadratic term structure models.

References

Ahn, Dong-Hyun, Robert F. Dittmar, and A. Ronald Gallant, 2001, Quadratic term structure models: Theory and evidence, *Review of Financial Studies*, forthcoming.

Akaike, Hirotugu, 1973, Information theory and an extension of the maximum likelihood principle, in B.N. Petrov and F. Csaki, eds., *Second International Symposium on Information Theory*, 267–281, Akademiai Kiado: Budapest.

Ang, Andrew, and Monika Piazzesi, 2002, A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables, Working Paper, Columbia University.

Backus, David K., Silverio Foresi, Christopher Telmer, 1999, Discrete-time models of bond pricing, in N. Jegadeesh and B. Tuckman, eds., *Advanced Fixed-Income Valuation Tools*, Wiley and Sons: New York, NY.

Backus, David K., Silverio Foresi, and Stanley E. Zin, 1998, Arbitrage opportunities in arbitrage-free models of bond pricing, *Journal of Business and Economic Statistics* 16, 13–26.

Bansal, Ravi, and S. Viswanathan, 1993, No arbitrage and arbitrage pricing: A new approach, *Journal of Finance* 48, 1231–1262.

Black, Fischer, Emanuel Derman, and William Toy, 1990, A one-factor model of interest rates and its application to Treasury bond options, *Financial Analysts Journal* 46, 33–39.

Black, Fischer, and Piotr Karasinski, 1991, Bond and option pricing when short rates are log-normal, *Financial Analysts Journal* 47, 52–59.

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

Boudoukh, Jacob, 1993, An equilibrium model of nominal bond prices with inflation-output correlation and stochastic volatility, *Journal of Money, Credit, and Banking* 25, 636-665.

Brandt, Michael W., and Kevin Q. Wang, 1999, Time-varying risk aversion and unexpected inflation, Working Paper, University of Pennsylvania.

Brown, Roger H., and Stephen M. Schaefer, 1994, The term structure of real interest rates and the Cox, Ingersoll, and Ross model, *Journal of Financial Economics* 35, 3–42.

Campbell, John Y., and John H. Cochrane, 1999, By force of habit: A consumption-based explanation of aggregate stock market behavior, *Journal of Political Economy* 107, 205–251.

Chapman, David A., 1997, Approximating the asset pricing kernel, *Journal of Finance* 52, 1383–1410.

Cochrane, John H., 2001, *Asset Pricing*, Princeton University Press: Princeton, NJ.

Cochrane, John H., and Monika Piazzesi, 2002, Bond risk premia, Working Paper, University of Chicago.

Constantinides, George M., 1990, Habit formation: A resolution of the equity premium puzzle, *Journal of Political Economy* 98, 519–543.

Cramér, Harald, 1925, On some classes of series used in mathematical statistics, *Proceedings of the Sixth Scandinavian Mathematical Congress*, 399-425.

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

Diebold, Francis X., and Canlin Li, 2002, Forecasting the term structure of government bond yields, Working Paper, University of Pennsylvania.

Dittmar, Robert F., 2002, Nonlinear pricing kernels, kurtosis preference, and the cross-section of equity returns, *Journal of Finance* 57, 369–403.

Duffee, Gregory R., 2002, Term premia and interest rate forecasts in affine models, Journal of Finance, Vol-57, 405-444.

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

Duffie, Darrell, Kenneth J. Singleton, 1993, Simulated moments estimation of Markov models of asset prices, *Econometrica* 50, 987–1007.

Dybvig, Philip H., 1989, Bond and Bond Option Pricing Based on the Current Term Structure, Working Paper, Washington University.

Foster, Dean P., and Daniel B. Nelson, 1996, Continuous record asymptotics for rolling sample variance estimators, *Econometrica* 64, 139–174.

Gallant, A. Ronald, and Douglas W. Nychka, 1987, Seminonparametric maximum likelihood estimation, *Econometrica* 55, 363–390.

Gallant, A. Ronald, and George Tauchen, 1989, Seminonparametric estimation of conditionally constrained heterogeneous processes: Asset pricing applications, *Econometrica* 57, 1091–1120.

Gallant, A. Ronald, and George Tauchen, 1992, A nonparametric approach to nonlinear time series analysis: Estimation and simulation, in D. Brillinger, P. Caines, J. Geweke, E. Parzen, M. Rosenblatt, and M. Taqqu, eds., *New Directions in Time Series Analysis, Part II*, 71–91, Springer-Verlag: New York, NY.

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

Hansen, Lars Peter, 1982, Large sample properties of generalized method of moments estimators, *Econometrica* 50, 1029–1053.

Harrison, J. Michael, and David M. Kreps, 1979, Martingales and arbitrage in multiperiod securities markets, *Journal of Economic Theory* 20, 381–408.

Heath, David, Robert A. Jarrow, and Andrew Morton, 1992, Bond pricing and the term structure of interest rates, *Econometrica* 60, 225–262.

Ho, Thomas S. Y., and Sang-Bin Lee, 1986, Term structure movement and pricing interest rate contingent claims, *Journal of Finance* 41, 1011–29.

Hull, John C., and Alan White, 1990, Pricing interest rate derivatite securities, *Review of Financial Studies* 3, 573–592.

Hull, John C., and Alan White, 1993, One-factor interest rate models and the valuation of interest rate derivative securities, *Journal of Financial and Quantitative Analysis* 28, 235–254.

Ingram, Beth Fisher, and Bong-Soo Lee, 1991, Simulation estimation of time-series models, *Journal of Econometrics* 47, 197–205.

Jamshidian, Farshid, 1991, Bond and option evaluation in the Gaussian interest rate model, in A. Chen, eds., *Research in Finance, Volume 9*, 131-170, JAI Press: Greenwich, CT.

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

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

McCulloch, J. Houston, and Heon-Chul Kwon, 1993, U.S. term structure data, 1947-91, Working Paper, Ohio State University.

McCulloch, J. Houston, 1971, Measuring the term structure of interest rates, *Journal of Business* 44, 19–31.

McCulloch, J. Houston, 1975, The tax adjusted yield curve, *Journal of Finance* 30, 811–829.

Powell, Mike J. D., 1981, *Approximation Theory and Methods*, Cambridge University Press: Cambridge, MA.

Schwartz, Stuart C., 1967, Estimation of probability density by an orthogonal series, *Annals of Mathematical Statistics* 38, 1261–1265.

Stambaugh, Robert F., 1988, The information in forward rates, *Journal of Financial Economics* 21, 41–70.

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

Table 1: Data

This tables describes annualized and continuously compounded yields to maturity of zero-coupon U.S. Treasury bonds from January 1959 through December 1998 (480 observations). The maturities range from one month to 10 years. Panels A and B present summary statistics for the yields and yield spreads relative to the one-month yield, respectively. Panels C and D describe the macroeconomic factors (consumption growth and inflation) and the yield-based factors (the level and conditional volatility of the one-month interest rate), respectively.

	Mean	Standard Deviation	Skewness	Kurtosis	Auto-Correlation
			Panel A: Yields		
1-Month	5.768	2.662	1.226	4.815	0.972
6-Month	6.316	2.749	1.205	4.568	0.979
12-Month	6.524	2.722	1.115	4.234	0.980
24-Month	6.777	2.647	1.024	3.956	0.983
60-Month	7.110	2.539	0.926	3.614	0.988
120-Month	7.318	2.482	0.756	3.219	0.992
			Panel B: Yield Spreads		
6-Month	0.548	0.430	1.352	7.809	0.575
12-Month	0.756	0.590	1.125	7.365	0.704
24-Month	1.009	0.800	0.307	5.023	0.805
60-Month	1.342	1.135	-0.161	3.581	0.883
120-Month	1.550	1.343	-0.177	3.114	0.908
			Panel C: Macroeconomic Factors		
Cons. Growth	1.010	0.682	-0.193	4.465	-0.218
Inflation	1.888	0.463	0.983	4.515	0.668
			Panel D: Yield-Based Factors		
Level	1.661	0.436	-0.310	2.869	0.976
Volatility	2.935	2.414	2.771	13.111	0.802

Table 2: In- and Out-of-Sample Performance with Macroeconomic Factors

This table reports the mean and mean absolute fitting and forecasting errors of first-, second-, and third-order log pricing kernels with consumption growth and inflation as factors and VAR or SNP factor dynamics. For each date t in the sample, we fit the pricing kernels to the cross-sections of bond prices from dates $t-\tau$ to t and then evaluate the pricing errors for the cross-section of bond prices at date t for panel A and at date $t+l$ for panel B, where l denotes the forecast horizon. $\tau=\infty$ denotes the fit to the whole panel of bond prices.

Panel A: In-Sample Fit

Order	Mean				Mean Absolute			
	$\tau=0$	$\tau=2$	$\tau=5$	$\tau=\infty$	$\tau=0$	$\tau=2$	$\tau=5$	$\tau=\infty$
				VAR Factor Dynamics				
1	0.024	0.031	0.059	0.073	0.113	0.230	0.390	0.899
2	0.014	0.010	0.005	-0.007	0.099	0.196	0.323	0.824
3	0.001	-0.006	-0.013	-0.022	0.075	0.179	0.309	0.809
				SNP Factor Dynamics				
1	-0.018	-0.021	-0.059	-0.073	0.106	0.202	0.345	0.853
2	-0.097	-0.006	-0.005	0.009	0.088	0.189	0.284	0.784
3	0.001	0.019	-0.026	-0.034	0.065	0.167	0.279	0.768

Panel B: Out-of-Sample Forecasts

	Forecast Horizon											
	3-Month			6-Month			9-Month			12-Month		
Order	$\tau=0$	$\tau=2$	$\tau=5$	$\tau=0$	$\tau=2$	$\tau=5$	$\tau=0$	$\tau=2$	$\tau=5$	$\tau=0$	$\tau=2$	$\tau=5$
						VAR Factor Dynamics						
1	0.594	0.514	0.475	0.753	0.763	0.680	0.881	0.820	0.815	1.001	0.954	0.950
2	0.556	0.467	0.420	0.716	0.672	0.619	0.855	0.785	0.778	0.925	0.875	0.879
3	0.544	0.445	0.414	0.686	0.651	0.588	0.834	0.760	0.754	0.907	0.862	0.868
						SNP Factor Dynamics						
1	0.502	0.471	0.442	0.671	0.647	0.647	0.819	0.774	0.781	0.990	0.893	0.899
2	0.497	0.417	0.383	0.651	0.563	0.566	0.775	0.677	0.681	0.889	0.795	0.819
3	0.489	0.405	0.373	0.641	0.557	0.550	0.743	0.652	0.666	0.875	0.773	0.787

Table 3: In- and Out-of-Sample Performance with Yield-Based Factors

This table reports the mean and mean absolute fitting and forecasting errors of first-, second-, and third-order log pricing kernels with the level and conditional volatility of the one-month interest rate as factors and SNP factor dynamics. For each date t in the sample, we fit the pricing kernels to the cross-sections of bond prices from dates $t-\tau$ to t and then evaluate the pricing errors for the cross-section of bond prices at date t for panel A and at date $t+l$ for panel B, where l denotes the forecast horizon. $\tau=\infty$ denotes the fit to the whole panel of bond prices.

Panel A: In-Sample Fit

	Mean				Mean Absolute			
Order	$\tau=0$	$\tau=2$	$\tau=5$	$\tau=\infty$	$\tau=0$	$\tau=2$	$\tau=5$	$\tau=\infty$
1	-0.053	-0.062	-0.080	-0.093	0.075	0.138	0.204	0.769
2	-0.032	-0.045	-0.063	-0.072	0.054	0.119	0.177	0.719
3	-0.010	-0.034	-0.040	-0.052	0.037	0.107	0.168	0.706

Panel B: Out-of-Sample Forecasts

|	Forecast Horizon											
	3-Month			6-Month			9-Month			12-Month		
Order	$\tau=0$	$\tau=2$	$\tau=5$	$\tau=0$	$\tau=2$	$\tau=5$	$\tau=0$	$\tau=2$	$\tau=5$	$\tau=0$	$\tau=2$	$\tau=5$
1	0.405	0.350	0.323	0.517	0.495	0.508	0.687	0.621	0.630	0.797	0.750	0.759
2	0.364	0.318	0.289	0.471	0.429	0.435	0.625	0.544	0.556	0.759	0.694	0.709
3	0.348	0.301	0.277	0.487	0.416	0.424	0.618	0.528	0.540	0.748	0.687	0.698

Figure 1: Term Structure of Interest Rates

This figures plots the panel of annualized and continuously compounded yields to maturity of zero-coupon U.S. Treasury bonds from January 1959 through December 1998 (480 observations). The maturities range from one month to ten years.

34
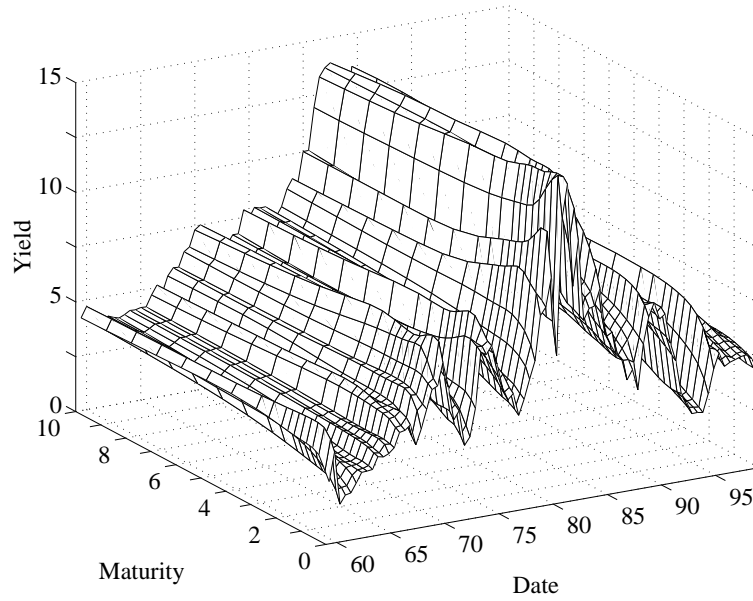
Figure 2: Mean Absolute Forecast Errors with Macroeconomic Factors

The first row of plots presents the out-of-sample mean absolute forecast errors for one-, five-, and ten-year bonds at time $t+l$ of pricing kernels that are fitted to the cross-section of bond prices at date t, where l denotes the forecast horizon. The factors are consumption growth and inflation with SNP factor dynamics. The circle, diamond, square symbols represent first-, second-, and third-order log pricing kernels, respectively. The second and third rows are for pricing kernels that are fitted to the cross-section of bond prices from date $t - \tau$ to date t, where $\tau = 2$ in the second row and $\tau = 5$ in the third row.
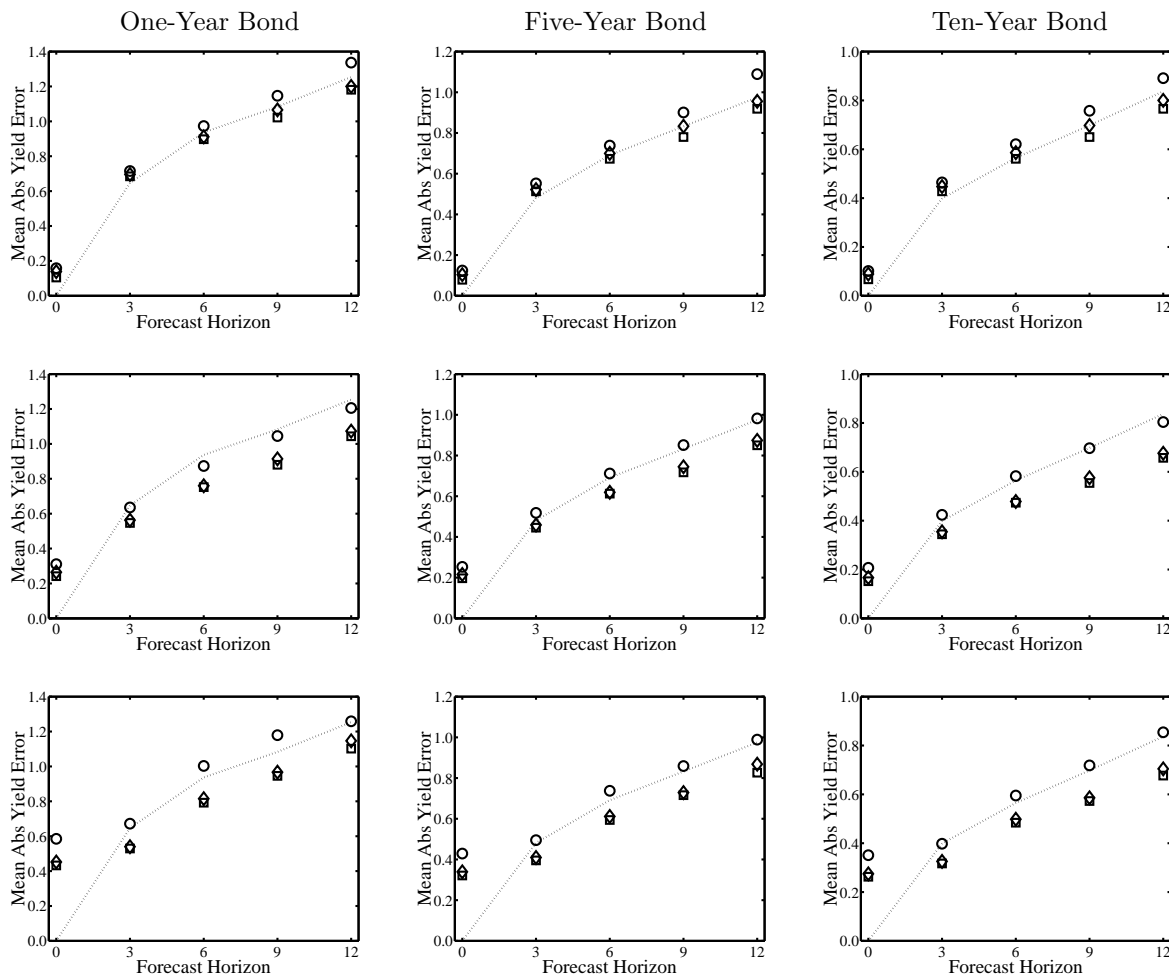
Figure 3: Mean Absolute Forecast Errors with Yield-Based Factors

The first row of plots presents the out-of-sample mean absolute forecast errors for one-, five-, and ten-year bonds at time $t + l$ of pricing kernels that are fitted to the cross-section of bond prices at date t, where l denotes the forecast horizon. The factors are the level and conditional volatility of the one-month interest rate with SNP factor dynamics. The circle, diamond, square symbols represent first-, second-, and third-order log pricing kernels, respectively. The second and third rows are for pricing kernels that are fitted to the cross-section of bond prices from date $t - \tau$ to date t, where $\tau = 2$ in the second row and $\tau = 5$ in the third row.