Available online at www.sciencedirect.com
SCIENCE @ DIRECT®

Journal of Econometrics 116 (2003) 147 – 180

JOURNAL OF Econometrics

www.elsevier.com/locate/econbase

Purebred or hybrid?: Reproducing the volatility in term structure dynamics

Dong-Hyun Ahn[a,b], Robert F. Dittmar[c], A. Ronald Gallant[d,*], Bin Gao[e]

[a] *College of Business Administration, Korea University, South Korea*
[b] *Department of Finance, Kenan-Flagler Business School, University of North Carolina, USA*
[c] *Department of Finance, Kelley School of Business, Indiana University, USA*
[d] *Department of Economics, University of North Carolina, CB 3305, Chapel Hill, NC 27599-3305, USA*
[e] *Department of Finance, Kenan-Flagler Business School, University of North Carolina, USA*

Abstract

This paper investigates the ability of mixtures of affine, quadratic, and non-linear models to track the volatility in the term structure of interest rates. Term structure dynamics appear to exhibit pronounced time varying or stochastic volatility. Ahn et al. (Rev. Financial Stud. xx (2001) xxx) provide evidence suggesting that term structure models incorporating a set of quadratic factors are better able to reproduce term structure dynamics than affine models, although neither class of models is able to fully capture term structure volatility. In this study, we combine affine, quadratic and non-linear factors in order to maximize the ability of a term structure model to generate heteroskedastic volatility. We show that this combination entails a tradeoff between specification of heteroskedastic volatility and correlations among the factors. By combining factors, we are able to gauge the cost of this tradeoff. Using efficient method of moments (Gallant and Tauchen, Econometric Theory 12 (1996) 657), we find that augmenting a quadratic model with a non-linear factor results in improvement in fit over a model comprised solely of quadratic factors when the model only has to confront first and second moment dynamics. When the full dynamics are confronted, this result reverses. Since the non-linear factor is characterized by stronger dependence of volatility on the level of the factor, we conclude that flexibility in the specification of both level dependence and correlation structure of the factors are important for describing term structure dynamics.

JEL classification: E43; C14; C15; C52

Keywords: Term structure models; Yield curve; Affine models; Quadratic models; Inverted square root models; Efficient method of moments; Simulation estimators

* Corresponding author.
E-mail address: ron.gallant@unc.edu (A.R. Gallant).

0304-4076/03/$ - see front matter
doi:10.1016/S0304-4076(03)00106-4

1. Introduction

Academic researchers and investment institutions have devoted a significant amount of their effort over the past two decades to developing and testing sophisticated models of the term structure of interest rates. These models are important both to academics and practitioners because the dynamics of interest rates have important implications for macroeconomic policy and the microeconomic decisions of agents in an economy. Studying this issue is particularly pressing in light of the drastic fluctuations in U.S. interest rates over the past three decades. These fluctuations demonstrate the need for a robust model that captures term structure dynamics.

An important issue in describing interest rate movements is modeling their volatility. Several empirical studies suggest that the volatility in the dynamics of interest rates are time-varying, and most likely stochastic (e.g. Longstaff and Schwartz, 1992; Engle and Ng, 1993; Gallant and Tauchen, 1998). This literature relies upon the premises that (i) term premia exhibit strong time variation and (ii) time variation in these premia are induced by time variation in interest rates and/or the state variables that govern their stochastic dynamics. These studies have been followed by a stream of research that investigates the empirical goodness-of-fit of alternative specifications of the stochastic differential equations of the *state variables* for the short rate, concentrating largely on explaining its time-varying volatility.[1] These studies represent an in-depth analysis of the performance of alternative models under the physical probability measure.

A smaller set of studies investigate the properties of alternative models under both the physical and risk-neutral probability measures, exploiting the information in both the time series and cross-section of bond yields. Because these studies gauge model performance based on both the physical measure and the equivalent martingale measure, they are able to empirically discriminate between models more rigorously. Representative studies in this vein include Dai and Singleton (2000) and Ahn et al. (2001). Dai and Singleton find that affine term structure models (hereafter referred to as *ATSM*s) are able to fit the *unconditional* term structure of volatility based on data from the late 1980s through the 1990s. Ahn et al. (2001) document that quadratic term structure models (*QTSM*s) empirically outperform *ATSM*s. However, the authors find that neither set of models can explain the stochastic features of the volatility of bond yields.

These empirical studies represent major progress in documenting the dynamics of the second moment of Treasury securities. However, despite this empirical progress, less effort has been made to connect the empirical dynamics of the second moment to a theoretical model. The empirical research suggests several dimensions along which existing models fail to fit the observed volatilities of yields to maturity. Ahn et al. (2001) observe that *ATSM*s generally produce a lower level of yield volatility and fall short in capturing its dynamic characteristics. *QTSM*s improve upon the *ATSM*s performance in both directions, but while the maximal *QTSM* model matches volatility dynamics fairly well, it fails to generate the high level of volatility observed in interest rates over different horizons. These findings suggest that, in order to understand the

[1] See Chan et al. (1992), Aït-Sahalia (1996a, b), Andersen and Lund (1997), and Gallant and Tauchen (1998) among many others for empirical investigations of the volatility of the short rate.

shortcomings of extant parametric models in volatility specification, we need to investigate existing models and their relative strengths and weaknesses by considering their implications for term structure volatility and its dynamics. In this paper, we conduct such an investigation.

Our starting point for investigating this issue is the set of equilibrium term structure models that have been developed over the past two decades.[2] This literature can be broadly separated into three alternative classes. The first and most popular class of term structure models are *ATSM*s, which designate an affine association between a set of underlying state variables and bond yields. This class includes the seminal works of Vasicek (1977) and Cox et al. (1985, CIR hereafter).[3] The class is generalized by Duffie and Kan (1996), who summarize the primitive assumptions underlying this set of models. Dai and Singleton (2000) characterize the admissibility of *ATSM*s and further contribute to the empirical literature by implementing *ATSM*s based on a theoretical analysis of the minimal conditions neccessary for identifying this class of models. Thus, the authors are able to specify restrictions on the Duffie and Kan framework that produce a maximally flexible and empirically identifiable *ATSM*. Following Dai and Singleton, denote an *ATSM* with m state variables with square-root processes (which can be potentially correlated) and $n-m$ Gaussian state variables as an $A_m(n)$. While the Gaussian factors are homoskedastic, the state variables with square-root process induce stochastic volatility with an order of $\frac{1}{2}$ as their name implies. This specification results in a "level" effect in the underlying state variables that generates stochastic volatility in the yields.

While the *ATSM* is able to generate stochastic volatility in yields, constraints imposed by its functional form limit its ability to generate stochastic volatility. First, only m of the n state variables may contribute to stochastic volatility of interest rates. This problem cannot be trivially solved by setting m to n because, as noted in Dai and Singleton (2000) and stressed in Ahn et al. (2001), while $m=n$ maximizes the *ATSM*s flexibility in specifying heteroskedastic volatility, it limits its flexibility in specifying conditional/unconditional correlations among state variables. Second, the dependence of volatility to level is restricted by an affine relationship between yields and interest rates. This issue may be important in light of studies (e.g. Aït-Sahalia, 1996b) that suggest a stronger order dependence of yield volatility on level.

The limitations in the specification of *ATSM*s are driven by its affine functional form. Ahn et al. (2001) demonstrate that these limitations can be overcome by

[2] We focus on equilibrium models in contrast to arbitrage models. The equilibrium approach is, as its name implies, an attempt to endogenize the term structure of interest rates by specifying an underlying economy using assumed preferences of either a representative agent or a stand-in aggregate household, his/her monetary and/or information constraints, and imposing market clearing. In contrast, the arbitrage approach relies upon extracting no arbitrage constraints from observed term structures and using this information to price contingent claims on either interest rate or fixed-income securities. Thus the latter approach is less suitable for interpreting model implications for the determination of the characteristics of the dynamics of the term structure of interest rates.

[3] These single-factor models are extended to multivariate versions by Langetieg (1980), Chen and Scott (1992), Longstaff and Schwartz (1992), Sun (1992), Pearson and Sun (1994), Balduzzi et al. (1996), Chen (1996), Andersen and Lund (1997), and Jegadeesh and Pennachi (1996).

eliminating the affine restriction. Compared to *ATSM*s, non-affine term structure models have evolved much more slowly. The representative non-affine term structure models are *QTSM*s. This class of models was first launched by Longstaff (1989) and pioneered by Beaglehole and Tenney (1991, 1992) and Constantinides (1992). Ahn et al. (2001) consolidate and systematically compare these models in a general framework by formally defining the *QTSM*s and building a maximally flexible *QTSM* which can be identified empirically.[4] The authors clarify the exact restrictions which reduce the general *QTSM* to existing sub-class models. The advantage of the *QTSM* relative to the *ATSM* in specifying heteroskedastic volatility comes from the inclusiveness of all state variables in generating the stochastic volatility. Ahn et al. (2001) also demonstrate that the *QTSM* is free of the trade-off between heteroskedastic volatility of interest rates and negative correlation among the state variables, while maintaining admissibility. However, the *QTSM* is isomorphic to the *ATSM* in its mechanism for generating volatility; the volatility of the interest rate is proportional to the level of the state variables. Although there is no strict one-to-one correspondence between the square root factors in the *ATSM* and the Gaussian factors in the *QTSM*, it is easily demonstrated that the contribution of an square root factor to volatility is equivalent to the contribution of the square of a Gaussian factor.[5] Thus in terms of generating volatility, $A_n(n)$ is equivalent to $Q(n)$.[6]

The last and least developed class of equilibrium term structure models are fully non-affine models, in which the state variables themselves evolve according to non-affine stochastic processes. The only model of which we are aware in this class is the inverted square-root model (*ISRM*) of Ahn and Gao (1999). This model is based on the notion that the interest rate is the inverse of a state variable which follows a square-root process. The framework is unique because the drift of the interest rate is a quadratic function of the underlying state variables and its volatility is governed by a power function of the state variable with an exponent of $\frac{3}{2}$. The advantage of this model is that it has the potential to magnify stochastic volatility by leveraging the interest rate in the volatility specification. However, one drawback of this model is that a closed-form solution (even in ordinary differential equation terms) does not exist once we incorporate non-trivial correlation among state variables.

From this starting set of equilibrium term structure models, we investigate whether a mixture of these three different classes of models can outperform purebred models in explaining the high volatility of interest rates in the US. Put differently, this paper explores the potential positive (or negative) synergy in combining heterogeneous parametric models, particularly in specifying the volatility of interest rates. If the hybrid model can improve upon the performance of parametric models in volatility specification, there might be an offsetting cost, which is of central importance to this paper.

[4] Leippold and Wu (2000) also characterize the *QTSM*s in a general framework but do not focus on empirical implementation.

[5] Using this concept, Ahn et al. (2001) formally prove that a particular version of the CIR model coincides with a constrained version of *QTSM* under certain relevant restrictions.

[6] However, this equivalence does not hold in a discrete horizon because of the feedback effect in the drift term of interest rates.

None of these questions have been addressed in the extant literature.[7] We denote a hybrid model with m square root factors, $n-m$ Gaussian factors, h quadratic factors, and k inverted-square root factors by $A_m(n)Q(h)I(k)$. Consequently the total number of state variables is $n+h+k$. With this purpose, we investigate different hybrid models and compare their overall performance in explaining the transition distribution of yields and in particular the fit of the volatility specification. We use the efficient method of moments (EMM) of Gallant and Tauchen (1996) to estimate a wide variety of $A_m(n)Q(h)I(k)$. Following Dai and Singleton (2000) and Ahn et al. (2001), we simultaneously use time series data on short- and long-term Treasury bond yields to explore the empirical properties of $A_m(n)Q(h)I(k)$. We focus primarily on hybrid models with $h=2$, namely, models with two quadratic factors augmented by one orthogonal affine (square-root or Gaussian) or inverted-square root factor.

We find several interesting results. First, when we focus on capturing just the first two conditional moments of the distribution of yields, we find that a model that combines quadratic and inverted square root factors outperforms the remaining models. We conclude that in this scenario, the loss in terms of modeling correlation among the state variables is outweighed by amplifying the effect of factor levels on yield volatility. The results suggest that, in modeling the conditional mean and volatility of interest rates, accounting for this strong level dependence is at least as important as flexibility in specifying correlations among the state variables. However, when we investigate a more complicated auxiliary model for the conditional density, we find that the pure *QTSM* model outperforms the hybrid Quadratic-Inverted-Square-Root model. This result suggests that the additional flexibility provided by the correlation structure of the *QTSM* is important in capturing shape deviations from conditional normality. In conjunction, the results suggest that a model that is able to incorporate correlations among the state variables and strong level dependence in volatility would fit the term structure data better.

The paper is organized as follows. In Section 2, we introduce the hybrid models, $A_m(n)Q(h)I(k)$, based on a combination of alternative purebred parametric models, *ATSM*s, *QTSM*s, and *ISRM*s. The following section provides a discussion of the term structure data and the EMM methodology that we use for examining the fit of the $A_m(n)Q(h)I(k)$. The empirical results of the EMM estimation coupled with their implication for heteroskedastic volatility are provided in Section 4, and Section 5 makes some concluding remarks and the implication of the paper for future studies.

2. Models

We define an economy represented by the augmented filtered probability space $(\Omega, \mathscr{F}_t, \mathbb{F}, \mathrm{P})$, where $\mathbb{F} = \{\mathscr{F}_t\}_{0 \leqslant t \leqslant \mathscr{T}}$. We assume the existence of a stochastic discount factor, $\pi(t)$, which defines the canonical valuation equation under the physical

[7] Strictly speaking, $A_m(n)$ itself can be viewed as a hybrid model of the CIR model and the Vasicek model. However, this class of hybrid models reside within the affine framework.

probability measure P:

$$x(t) = \mathrm{E}_t^{\mathrm{P}}\left[\frac{\pi(T)}{\pi(t)}x(T)\right],$$

where $x(t)$ is the price of an asset, with $x(t,\omega):[0,\infty)\times\Omega\to R^+$. It is well known that alternative asset pricing models differ in their description of the time-series evolution of the stochastic discount factor or pricing kernel, $\pi(t)$. The absence of arbitrage opportunities coupled with the assumption of a complete market ensures the existence of a unique equivalent probability measure, Q, under which a pure discount bond price is given by

$$P_B(t,T) = \mathrm{E}_t^{Q}\left[\exp\left(-\int_t^T r(s)\,\mathrm{d}s\right)\right]. \tag{1}$$

where $r(t)$ is the instantaneous risk free interest rate. The change of measure is dictated by the Radon-Nikodym derivative

$$\xi(t,T) \triangleq \frac{\mathrm{d}Q(t,T)}{\mathrm{d}P(t,T)} = \exp\left(\int_t^T \varphi(s)\,\mathrm{d}w_q(s) - \frac{1}{2}\int_t^T \varphi(s)^2\,\mathrm{d}s\right),$$

where $\varphi(s)$ denotes a vector of diffusion parameters in $\xi(t,T)$ and $w_q(t)$ denotes a vector of independent standard Brownian motions. The uniqueness of the pricing kernel guarantees that

$$\frac{\pi(T)}{\pi(t)} = \xi(t,T)\exp\left(-\int_t^T r(s)\,\mathrm{d}s\right).$$

From the bond valuation equation (1), it is clear that only the factor dynamics under the Q measure have an impact on bond prices. However, to completely specify a term structure model with both time series and cross-sectional implications, we need to specify ξ as well. To see this, we express the stochastic differential equation of the pricing kernel as follows:

$$\frac{\mathrm{d}\pi(t)}{\pi(t)} = -r(s)\,\mathrm{d}t + \mathrm{d}\xi(t,t+) = -r(s)\,\mathrm{d}t + \varphi(t,t+)\,\mathrm{d}w_q(t).$$

We claim that three assumptions completely specify a term structure model:

(A1) *The relationship between the interest rate*, $r(t)$ *and the underlying state variables*, $Y(t)$
(A2) *The stochastic differential equations of the state variables*, $\mathrm{d}Y(t)$
(A3) *The diffusion process of the stochastic discount factor*, $\varphi(t)$

Note that (A1) determines the drift of the stochastic discount factor, (A2) dictates the stochastic evolution of the drift of the stochastic discount factor, and (A3) governs its diffusive evolution or equivalently the market price of factor risks since the factor risk premium is $-\mathrm{cov}_t(\mathrm{d}\pi(t)/\pi(t),\mathrm{d}Y(t)) = -\mathrm{cov}_t(\varphi(t)\,\mathrm{d}w_q(t),\mathrm{d}Y(t))$ as shown by Merton (1973) and Cox et al. (1985). To construct a term structure model, we have discretion over the choice of (A1)–(A3) subject to (i) the regularity conditions for the existence of solutions to each stochastic differential equation, and (ii) admissibility of the stochastic discount factor.[8]

In this study, we investigate hybrid models based on three representative term structure models: *ATSM*, *QTSM*, and *ISRM* from the viewpoint of their underlying assumptions regarding (A1)–(A3). Without loss of generality, we make a simplifying assumption about (A3) by directly specifying the market price of factor risks, $\lambda = -\mathrm{cov}_t(\mathrm{d}\pi(t)/\pi(t), \mathrm{d}Y(t))$. By assuming that N unobserved state variables $Y(t) = (Y_1(t), Y_2(t), \ldots, Y_N(t))$ follow various stochastic processes, we then investigate the relationship between pure discount bond prices and factors under different model specifications. In addition, we analyze the *canonical* representation of term structure models which lend themselves to empirical implementation in the presence of the unobservable state variables.

The hybrid model is based on the following assumptions:

A1. *The nominal instantaneous interest rate is a function of the underlying state variables*:

$$r(t) = \delta_0 + \delta_A' Y_A(t) + Y_I(t)' 1_k + Y_Q(t)' \Psi Y_Q(t) \tag{2}$$

where δ_0 *is a constant*; $\delta_A > 0_n$; $Dim(Y_A) = n$, $Dim(Y_Q) = h$, *and* $Dim(Y_I) = k$. *We set the total number of state variables* $N = n + h + k$. Ψ *is a symmetric h-dimensional matrix with unit diagonal elements. A, I, and Q represent affine, inverted square root, and quadratic factors, respectively.*

A2. *The stochastic differential equations for each set of state variables are identical to the specifications of those in Dai and Singleton* (2000), *Ahn et al.* (2001), *and Ahn and Gao* (1999).[9]

A3. *The market prices of factor risk are also identical to the specification of those in each of the respective papers above.*

Thus defined, the instantaneous interest rate follows a hybrid process of affine, quadratic and inverted square root factors. One crucial assumption embedded in A2 is that the set of state variables in different set of models are orthogonal to each other. That is,

$$\langle \mathrm{d}Y_i, \mathrm{d}Y_j \rangle = 0 \quad \forall i \neq j, \text{ where } i, j = A, Q, I.$$

These assumptions are necessary for us to have an analytical representation for the bond prices. In some cases, the inter-orthogonality conditions are necessary for identification reasons. For example, in the presence of unobservable state variables, a combination of the Gaussian *ATSM*s and the *QTSM*s requires orthogonality among the *ATSM* state variables and the *QTSM* state variables for an empirical identification of the model.[10] It is true that we may still have an admissible term structure model with non-trivial cross-model correlations, however, the model will not be solvable in closed-form.

[8] A sufficient condition for (i) is that both drift and diffusion coefficients satisfy the Lipschitz and growth conditions. A sufficient condition for (ii) is the satisfaction of Novikov condition.

[9] For details, see Dai and Singleton (2000) for *ATSM*s, Ahn et al. (2001) for *QTSM*s and Ahn and Gao (1999) for *ISRM*s.

[10] Please contact authors for the proof.

Given the specification of the interest rate structure (2), and the independent nature of factors from different model families, it is easy to show that the price of a pure discount bond is given by

$$P_B(t,\tau) = P_A(t,\tau)P_Q(t,\tau)P_I(t,\tau)$$

where τ is the bond maturity. $P_A(t,\tau)$, $P_Q(t,\tau)$, and $P_I(t,\tau)$ are defined by the following equations,

$$P_A(t,\tau) = \exp[\gamma_{0A}(\tau) + \gamma_{1A}(\tau)' Y_A(t)], \tag{3}$$

$$P_Q(t,\tau) = \exp[\gamma_{0Q}(\tau) + \gamma_{1Q}(\tau)' Y_Q(t) + Y_Q(t)'\gamma_{2Q}(\tau) Y_Q(t)], \tag{4}$$

$$P_I(t,\tau) = \Pi_{i=1}^{k} \frac{\Gamma(\eta_i - \varphi_i)}{\Gamma(\eta_i)} M(\varphi_i, \eta_i, -x(Y_{Ii}(t), t, \tau)) x(Y_{Ii}(t), t, \tau)^{\varphi_i}. \tag{5}$$

where

$$M(a,b,y) = \frac{\Gamma(b)}{\Gamma(b-a)\Gamma(a)} \int_0^1 \mathrm{e}^{yz} z^{a-1}(1-z)^{b-a-1}\,\mathrm{d}z,$$

and

$$x(Y_{Ii}(t), t, \tau) = \frac{2b_i}{\sigma_{ii}^2[\exp(b_i\tau) - 1] Y_{Ii}(t)},$$

with γ_A, and γ_Q being governed by two sets of ordinary differential equations, b_i, A_{ii}, and σ_{ii}, being constants, and η_i, φ_i, and ϕ_i being functions of these constants. [11]

In the following subsections, we specify each of the models that we investigate and discuss their implications for heteroskedastic volatility. As in Dai and Singleton (2000) and Ahn et al. (2001), we focus only on three factor models, but our primary interest is in the performance of models with different factor combinations. There are two issues that must be considered in constructing hybrid models in order to maintain both tractability and flexibility:

- Tractability: Dai and Singleton (2000) emphasize that correlation among factors is an important ingredient in describing bond price dynamics. Consequently, an ideal framework is driven by three correlated factors. However, analytical solutions to multivariate correlated cases are available only for *ATSM*s and *QTSM*s. In order to ensure tractability, we assume that factors of different types (affine, quadratic, and inverted square root) are independent.
- Correlation Structure: As noted above, Dai and Singleton (2000) suggest that it is important to have maximal flexibility in the specification of correlated factors in order to generate the correlations observed in bond yields. In order to maximize this flexibility, we examine models that have at least two correlated factors of the same type. In particular, we emphasize models with two quadratic factors augmented by an

[11] For details, see Dai and Singleton (2000) about *ATSM*s, and Ahn et al. (2001) about *QTSM*s. Ahn and Gao (1999) presents the bond price for single factor *ISRM*. We extend their model to an independent multifactor solution. Details available upon request from the authors.

additional independent factor of type either $ATSM$ or $ISRM$ and contrast our results with those of maximally flexible quadratic and/or affine models.

2.1. Two quadratic and one affine (square root) model: $A_1(1)Q(2)I(0)$

With two quadratic and one square-root factor, this model is expected to underperform the quadratic model. In terms of volatility specification, this model has the same order of volatility contribution to the three-factor $QTSM$. However, the loss of correlation structure between the square root factor and the quadratic factors will hurt the model's performance. However, this model provides some insight into the actual cost of weakening the correlation structure among the state variables.

The underlying factor process is represented by two quadratic factors and one square-root factor under the equivalent martingale measure Q:

$$\mathrm{d}Y(t) = \left[\begin{pmatrix} b_1 \\ b_2 \\ \tilde{\kappa}\tilde{\theta} \end{pmatrix} + \begin{pmatrix} A_{11} & 0 & 0 \\ A_{21} & A_{22} & 0 \\ 0 & 0 & -\tilde{\kappa} \end{pmatrix} Y(t)\right] \mathrm{d}t + \begin{pmatrix} \sigma_{11} & 0 & 0 \\ 0 & \sigma_{22} & 0 \\ 0 & 0 & \sigma\sqrt{Y_3(t)} \end{pmatrix} \mathrm{d}\tilde{Z}(t)$$

and under the physical process

$$\mathrm{d}Y(t) = \left[\begin{pmatrix} \Phi_{0,1} \\ \Phi_{0,2} \\ \kappa\theta \end{pmatrix} + \begin{pmatrix} \Phi_{1,11} & 0 & 0 \\ \Phi_{1,21} & \Phi_{1,22} & 0 \\ 0 & 0 & -\kappa \end{pmatrix} Y(t)\right] \mathrm{d}t + \begin{pmatrix} \sigma_{11} & 0 & 0 \\ 0 & \sigma_{22} & 0 \\ 0 & 0 & \sigma\sqrt{Y_3(t)} \end{pmatrix} \mathrm{d}Z(t)$$

with prices given by

$$P_B(t,\tau) = P_Q(t,\tau)P_A(t,\tau)$$

where P_Q represents the price for a two-factor quadratic bond price and P_A represents the price for a one-factor Square-Root bond price presented in (4) and (4) with $m=n=1$, and $h=2$. The parameters under Q and P measures are linked by $\tilde{\kappa}\tilde{\theta}=\kappa\theta$, and $\tilde{\kappa}=\kappa-\lambda_1$.

2.2. Two quadratic and one inverted square-root model: $A_0(0)Q(2)I(1)$

Of the models we consider, this one offers potential trade-off between correlation and level dependence. By replacing a correlated quadratic factor with an independent inverted-square-root factor, we lose flexibility in correlation structure. However, as strongly evidenced by Chan et al. (1992), Aït-Sahalia (1996a, b), and Ahn and Gao (1999), interest rate volatility increases at an increasing rate in interest rate levels with a level dependence on the order of approximately 1.5. The inverted-square-root factor exhibits this stronger level dependence (the order of the diffusion term is 1.5), and thus may potentially result in better goodness-of-fit.

The model is based on two quadratic factors and one inverted square-root factor following the risk-neutral process

$$\mathrm{d}Y(t)=\left[\begin{pmatrix} b_1 \\ b_2 \\ \tilde{\kappa}\tilde{\theta}Y_3(t)\end{pmatrix}+\begin{pmatrix} A_{11} & 0 & 0 \\ A_{21} & A_{22} & 0 \\ 0 & 0 & -\tilde{\kappa}Y_3(t)\end{pmatrix}Y(t)\right]\mathrm{d}t$$

$$+\begin{pmatrix}\sigma_{11} & 0 & 0 \\ 0 & \sigma_{22} & 0 \\ 0 & 0 & \sigma\sqrt{Y_3(t)^3}\end{pmatrix}\mathrm{d}\tilde{Z}(t)$$

and the physical process

$$\mathrm{d}Y(t)=\left[\begin{pmatrix} \Phi_{0,1} \\ \Phi_{0,2} \\ \kappa\theta Y_3(t)\end{pmatrix}+\begin{pmatrix} \Phi_{1,11} & 0 & 0 \\ \Phi_{1,21} & \Phi_{1,22} & 0 \\ 0 & 0 & -\kappa Y_3(t)\end{pmatrix}Y(t)\right]\mathrm{d}t$$

$$+\begin{pmatrix}\sigma_{11} & 0 & 0 \\ 0 & \sigma_{22} & 0 \\ 0 & 0 & \sigma\sqrt{Y_3(t)^3}\end{pmatrix}\mathrm{d}Z(t)$$

with prices given by

$$P_B(t,\tau)=P_Q(t,\tau)P_I(t,\tau)$$

where P_Q represents the price for a two-factor quadratic bond price and P_I represents the price for a one-factor Inverted-Square-Root bond price presented in (4) and (5) with $h=2$, and $k=1$. The restrictions imposed by the market price of risk manifests themselves as $\tilde{\kappa}\tilde{\theta}=\kappa\theta$, and $\tilde{\kappa}=\kappa-\lambda_1$.

For the completeness of model specifications, we also present the two benchmark models as special cases of hybrid models.

2.3. *Hybrid ATSMs:* $A_2(3)Q(0)I(0)$

Following Dai and Singleton (2000), we choose $A_2(3)$ as the benchmark case. In this case, the underlying factor process is governed by two Square-Root factors and one Gaussian factor following the risk-neutral process

$$dY(t) = \begin{pmatrix} \tilde{\kappa}_{11} & 0 & 0 \\ \tilde{\kappa}_{21} & \tilde{\kappa}_{22} & 0 \\ 0 & 0 & \tilde{\kappa}_{33} \end{pmatrix} \left[\begin{pmatrix} \tilde{\theta}_1 \\ \tilde{\theta}_2 \\ 0 \end{pmatrix} - Y(t) \right] dt + \begin{pmatrix} \sqrt{\beta_{11} Y_1(t)} & 0 & 0 \\ 0 & \sqrt{\beta_{22} Y_1(t)} & 0 \\ \sigma_{31}\sqrt{\beta_{11} Y_1(t)} & 0 & \sqrt{Y_1(t)} \end{pmatrix} d\tilde{Z}(t)$$

and the physical process

$$dY(t) = \begin{pmatrix} \kappa_{11} & 0 & 0 \\ \kappa_{21} & \kappa_{22} & 0 \\ 0 & 0 & \kappa_{33} \end{pmatrix} \left[\begin{pmatrix} \theta_1 \\ \theta_2 \\ 0 \end{pmatrix} - Y(t) \right] dt + \begin{pmatrix} \sqrt{\beta_{11} Y_1(t)} & 0 & 0 \\ 0 & \sqrt{\beta_{22} Y_1(t)} & 0 \\ \sigma_{31}\sqrt{\beta_{11} Y_1(t)} & 0 & \sqrt{Y_1(t)} \end{pmatrix} dZ(t)$$

with prices given by

$$P_B(t,\tau) = P_A(t,\tau)$$

and $P_A(t,\tau)$ given by (3) when $m = 2$, and $n = 3$. The restrictions by the market price of risk are $\tilde{\kappa}\tilde{\theta} = \kappa\theta$, and $\tilde{\kappa} = \kappa - \lambda$, where λ is a 3×3 matrix with λ_{11}, λ_{22}, λ_{33} as diagonal elements and zero otherwise.

2.4. Quadratic models: $A_0(0)Q(3)I(0)$

*QTSM*s are the representative non-affine class models. These models are attributable to Longstaff (1989), Beaglehole and Tenney (1991, 1992), Constantinides (1992). The maximally flexible *QTSM*s which lend themselves to identification are developed by Ahn et al. (2001).

The underlying factor process is determined by two quadratic factors and one Square-Root factor following the risk-neutral process

$$dY(t) = \left[\begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} + \begin{pmatrix} A_{11} & 0 & 0 \\ A_{21} & A_{22} & 0 \\ A_{31} & A_{32} & A_{33} \end{pmatrix} Y(t) \right] dt + \begin{pmatrix} \sigma_{11} & 0 & 0 \\ 0 & \sigma_{22} & 0 \\ 0 & 0 & \sigma_{33} \end{pmatrix} d\tilde{Z}(t)$$

and the physical process

$$dY(t) = \left[\begin{pmatrix} \Phi_{0,1} \\ \Phi_{0,2} \\ \Phi_{0,3} \end{pmatrix} + \begin{pmatrix} \Phi_{1,11} & 0 & 0 \\ \Phi_{1,21} & \Phi_{1,22} & 0 \\ \Phi_{1,31} & \Phi_{1,32} & \Phi_{1,33} \end{pmatrix} Y(t) \right] dt$$

$$+\begin{pmatrix} \sigma_{11} & 0 & 0 \\ 0 & \sigma_{22} & 0 \\ 0 & 0 & \sigma_{33} \end{pmatrix} \mathrm{d}Z(t)$$

with prices given by

$$P_B(t,\tau) = P_Q(t,\tau).$$

and $P_Q(t,\tau)$ given by (4) when $h = 3$.

In our empirical implementation we emphasize the two hybrid models, $A_1(1)Q(2)I(0)$ and $A_0(0)Q(2)I(1)$. These two models are similar in that they sacrifice some ability to model correlations among the state variables. However, the incorporation of the inverted square root factor in $A_0(0)Q(2)I(1)$ allows us to capture greater level dependence of volatility on the factor. Thus, comparing these two models shows us just what benefit (or cost) is provided by using a higher order level dependence in volatility. This benefit or cost is isolated from the cost of flexibility in correlation specification that arises from comparison with a pure Quadratic or hybrid Square-Root/Gaussian model. [12]

3. Data and methods

3.1. Term structure data

In order to investigate the implications of the models for the term structure of interest rates, we combine two sets of data. The first data set is the McCulloch and Kwon (1993) zero coupon yields, which cover the period January, 1952 through February, 1991, and are sampled at a monthly frequency. The second data set used is zero-coupon yield data covering the period November, 1971, through December, 1999. These data are formed using the methods in Bliss (1997). [13] For the overlap period of November, 1971, through February, 1991, the data were combined by using a convex combination

$$\lambda * (\text{McCulloch-Kwon data}) + (1-\lambda) * (\text{Waggoner data})$$

where λ declines linearly from one in October, 1971, to zero in March, 1991. The data used in the analysis and shown in the plots are from 1953 onward; initial lags for estimation are taken from the 1952 data.

For the purposes of the analysis of the model's ability to fit the term structure of interest rates, we utilize three yields; the 6-month T-Bill yield and the 3-year and 10-year bond yield. These maturities are similar to those examined in comparable studies, e.g. Dai and Singleton (2000), Ahn et al. (2001). As these yields cover short-, intermediate-, and long-term bonds, we feel that they provide a reasonable description of the term structure of interest rates at a given point in time. The data are plotted in Fig. 1, which shows that the sample period covers a wide range of interest rate

[12] Along with the aforementioned hybrid models, we implemented many other hybrid models. The empirical performance of these models is not as good as the models reported. We do not report the description of these models and their performance for brevity of the paper.

[13] Thanks to Daniel Waggoner for making these data available.

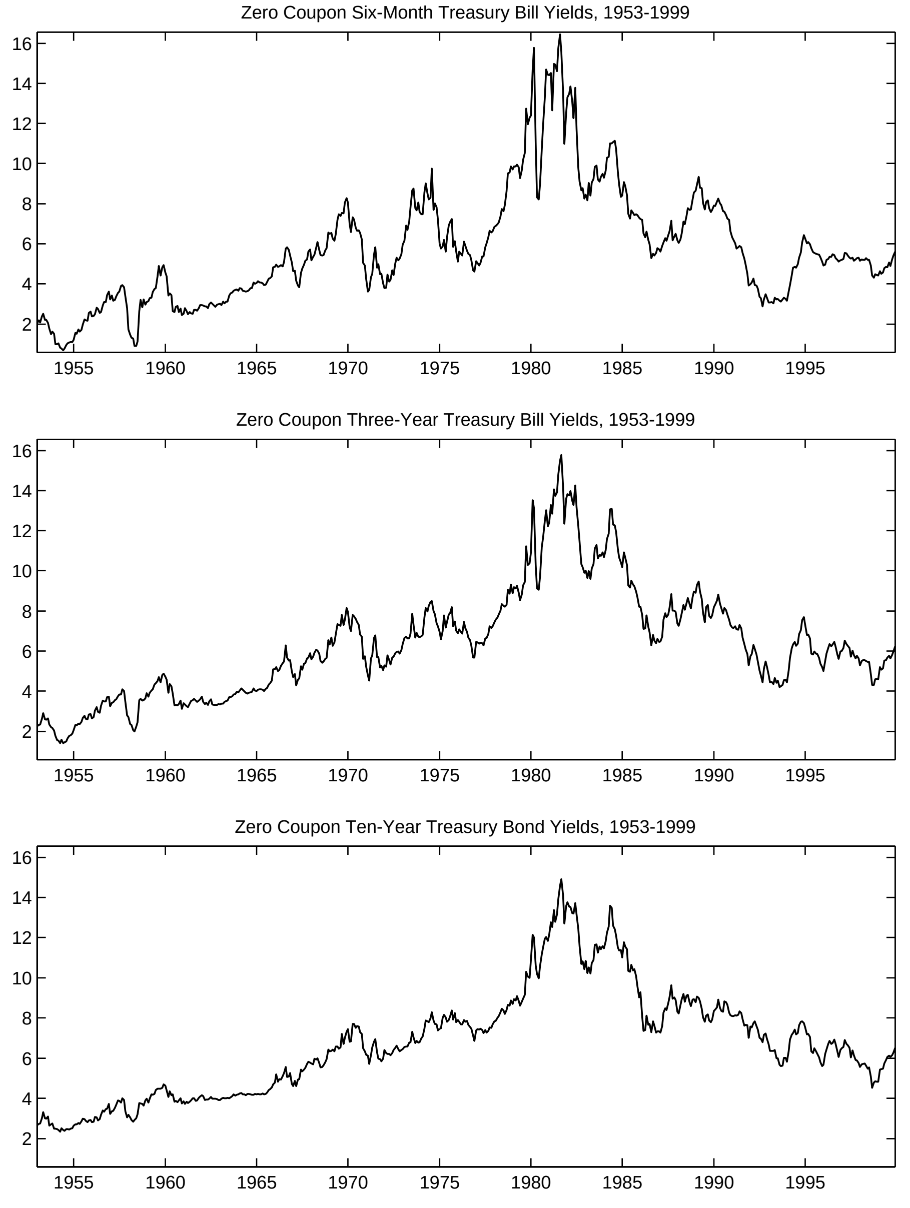

Fig. 1. Zero coupon yield data. The plots present the zero coupon yield data used in the study. The data are a combination of the data from McCulloch and Kwon (1993) and data produced using the methods in Bliss (1997).

regimes, from very low levels in the 1950s to the high rate regime of the early 1980s. Thus, the sample captures periods of relative stability in interest rates as well as periods punctuated by high volatility.

3.2. The efficient method of moments

One of the features of the term structure models investigated in this paper is that the state variables in the models are latent. As a result, estimation of the parameters of the model is complicated by the need to estimate the parameters of an unobserved stochastic process. Furthermore, since the model is expressed in continuous time, it is necessary to avoid issues of discretization bias (Aït-Sahalia, 1996a, b). Recent econometric advances have allowed researchers to address both of these issues through the use of simulated method of moments techniques. The simulated method of moments procedure that we employ here is efficient method of moments (EMM).

The theory of EMM estimation is developed in Gallant and Tauchen (1996) and is extended to non-Markovian data with latent variables in Gallant and Long (1997). An expository discussion of the method is in Gallant and Tauchen (2001). Briefly the ideas are as follows.

Suppose that $f(y_t|x_{t-1},\theta)$ is a reduced form model for the discretely sampled data, where x_{t-1} is the state vector of the observable process at time $t-1$ and y_t is the observable process, which is a vector of three bond yields in our application. An example of such a reduced form model is a GARCH(1,1), which is actually one of our choices below. If this reduced form model, which we shall call a score generator, is fitted by maximum likelihood to get an estimate $\tilde{\theta}_n$, then the average of the score over the data $\{\tilde{y}_t\}_{t=1}^n$ satisfies

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\partial}{\partial\theta}\log f(y_t|x_{t-1},\tilde{\theta}_n)=0$$

because these are the first order conditions of the optimization problem. If $p(y_t|x_{t-1},\rho)$ represents a structural model with parameter vector ρ, such as our model $A_2(3)Q(0)I(0)$ with $\rho=(\kappa_{11},\kappa_{21},\kappa_{22},\kappa_{33},\theta_1,\theta_2,\beta_{11},\beta_{22},\sigma_{31},\lambda_{11},\lambda_{22},\lambda_{33})$, then one would expect that a similar average over a long simulation $\{\hat{y}_t\}_{t=1}^N$ from the structural modal, namely

$$m(\rho,\theta)=\frac{1}{N}\sum_{t=1}^{N}\frac{\partial}{\partial\theta}\log f(\hat{y}_t|\hat{x}_{t-1},\theta),$$

would satisfy $m(\rho^o,\tilde{\theta}_n)=0$, where ρ^o denotes the true but unknown value of ρ.[14] One can try to solve $m(\rho,\tilde{\theta}_n)=0$ to get an estimate $\hat{\rho}_n$ of the parameter vector of the structural model. In general this cannot be done because the dimension of θ is larger than the dimension of ρ in most applications. To compensate for this, one estimates ρ by $\hat{\rho}_n$ that minimizes the GMM criterion

$$m'(\rho,\tilde{\theta}_n)(\tilde{\mathscr{I}}_n)^{-1}m(\rho,\tilde{\theta}_n)$$

with weighting matrix

$$\tilde{\mathscr{I}}_n=\frac{1}{n}\sum_{t=1}^{n}\left[\frac{\partial}{\partial\theta}\log f(\tilde{y}_t\,|\,\tilde{x}_{t-1},\tilde{\theta}_n)\right]\left[\frac{\partial}{\partial\theta}\log f(\tilde{y}_t\,|\,\tilde{x}_{t-1},\tilde{\theta}_n)\right]'.$$

[14] This condition will hold exactly in the limit as N and n tend to infinity under the standard regularity conditions of quasi maximum likelihood.

This choice of weighting matrix presupposes that the score generator fits the data well. If not, then one of the more complicated weighting matrices in Gallant and Tauchen (1996) should be considered. The estimator $\hat{\rho}_n$ is asymptotically normal.

If the structural model is correctly specified, then the statistic

$$L_0 = nm'(\hat{\rho}_n, \tilde{\theta}_n)(\tilde{\mathscr{I}}_n)^{-1}m(\hat{\rho}_n, \tilde{\theta}_n)$$

has the chi-squared distribution on $\dim(\theta) - \dim(\rho)$ degrees freedom. This is the familiar test of overidentifying restrictions in GMM nomenclature and is used to test model adequacy. A chi-squared is asymptotically normally distributed as degrees freedom increase. Therefore, for ease of interpretation, the statistic L_0 is often redundantly reported as a z-statistic, as we do in our tables.

The vector $m(\hat{\rho}_n, \tilde{\theta}_n)$ can be normalized by its standard error to get a vector of t-statistics. These t-statistics can be interpreted much as normalized regression residuals. They are often very informative but are subject to the same risk as the interpretation of regression residual; namely, a failure to fit one characteristic of the data can show up not at the score of the parameters that describe that characteristic but elsewhere due to correlation. Nonetheless, as with regression residuals, inspecting normalized $m(\hat{\rho}_n, \tilde{\theta}_n)$ is usually the most informative diagnostic available.

If the score generator is a poor fit to the data or the chi-squared test of model adequacy L_0 is not passed, then the analysis must be viewed as a calibration exercise rather than classical statistical inference. One might, for instance, deliberately choose a score generator that represents only some characteristics of the data to study the ability of a structural model to represent only those characteristics. We do this below. One might also use a rejected model to price options, arguing that it is the best available even though it was rejected. The use of EMM for calibration is discussed in Gallant et al. (1999).

Gallant and Tauchen (2001) recommend the seminonparametric (SNP) model as a general purpose score generator and that is our choice here. The SNP model is a vector autoregression (VAR) on L_u lags with a GARCH(L_g, L_r) conditional variance. The innovation density is a Hermite density of degree K_z. A Hermite density has the form of a polynomial times the multivariate standard normal density, which product then normalized to integrate to one. Too allow for conditional heterogeneity over and above that allowed by GARCH, the coefficients of the polynomial in the Hermite density are themselves polynomials of degree K_x in L_p lags of the data. Because the number of terms in a polynomial expansion become exponentially large as the dimension increases, two additional tuning parameters are introduced: $I_z > 0$ implies that all interactions larger than $K_z - I_z$ are suppressed; similarly for I_x. The tuning parameters of an SNP model are, therefore, $(L_u, L_g, L_r, L_p, K_z, I_z, K_x, I_z)$ and they are selected by following an upward model expansion path, using the BIC criterion to select the best model along the path.

The auxiliary model can be viewed as a summary of the data. It is accomplished by, in effect, projecting the data onto the SNP model and is therefore called the projection step of an EMM investigation. Extraction of structural parameters from the summary by minimizing the chi squared criterion is called the estimation step. In a later section, we shall describe a third step, reprojection, that often accompanies an EMM investigation.

With respect to our data, i.e. the McCulloch-Kwon-Waggoner data, among other things, we want to explore the ability of hybrid models to represent the first and second conditional moments of the data while ignoring other characteristics, such as a fat-tailed innovation density. For this, we fit a GARCH(1,1) score to the data as implemented by the $(L_u, L_g, Lr, L_p, K_z, I_z, K_x, I_z) = (11110000)$ SNP specification. [15] Following the upward BIC selection protocol, $(L_u, L_g, Lr, L_p, K_z, I_z, K_x, I_z) = (11114300)$ is selected. To compare, Dai and Singleton (2000) select a 10214000 score for y_t comprised of the six-month LIBOR, two-year swap yield, and ten-year swap yield over the period 1987–1996; Ahn et al. (2001) select a 10414300 score for the McCulloch-Kwon data at 3-month, 12-month, and ten-year maturities over the period 1946–1991. We expect inferences with respect to the 11114300 score to have classical statistical validity. Inferences with respect to the 11110000 score must be viewed as a calibration exercise.

To compute $m(\rho, \theta)$, we set the simulation length to $N = 50,000$. We repeated the computations for the 11114300 score with a simulation length T=100,000 to check for robustness. The results were materially unchanged: the parameter estimates remain virtually the same and the z-statistics increase by no more than 1.0. The public domain EMM package that we use, which includes SNP, is available at ftp.econ.duke.edu in directory pub/get/emm.

4. Estimation results

In this section we discuss various diagnostics for the set of term structure models investigated in the paper. In particular, we focus on specification testing using the EMM procedure (Gallant and Tauchen, 1996) and qualitative analysis using the reprojection methodology of Gallant and Tauchen (1998). We first focus on the simpler SNP specification to isolate the implications of the term structure models for conditional means and volatilities. This must be viewed as a calibration exercise. We then proceed to investigate the Schwartz-preferred specification in order to gauge the impact of deviations from conditional normality. This may be viewed as asymptotically justified classical statistical inference.

4.1. EMM specification tests

4.1.1. Results with the 11110000 score

In this subsection, we investigate the simpler 11110000 score generator that incorporates only VAR and GARCH effects. As mentioned above, we investigate this score because of an interest in the ability of hybrid models to fit the conditional mean and volatility characteristics of the data, as the title of the paper suggests. Because the conditional first and second moments completely characterize the conditional normal density, the assumption of normality focuses attention on those conditional characteristics alone.

[15] Due to a quirk in the public domain SNP code that we use, L_p must be positive even though $K_x = 0$.

Table 1
Specification tests of term structure models 11110000 Score, $N = 50000$

Coefficient	$A_0(0)Q(3)I(0)$	$A_1(1)Q(2)I(0)$	$A_0(0)Q(2)I(1)$	Coefficient	$A_2(3)Q(0)I(0)$
δ_0	0.0409 (0.0136)	0.0663 (0.0396)	0.0652 (0.0094)	κ_{11}	0.0148 (0.0019)
Ψ_{21}	−0.4732 (1.7933)	−8.7812 (12.2071)	−9.2078 (0.5624)	κ_{21}	0.1069 (0.1284)
Ψ_{31}	0.2409 (1.6110)			κ_{22}	0.2876 (0.0508)
Ψ_{32}	−0.9395 (0.1227)			κ_{33}	4.6249 (0.8871)
b_1	−0.0907 (0.2696)	0.0536 (0.0076)	−0.0035 (0.0000)	θ_1	0.0061 (0.0016)
b_2	0.0551 (0.0080)	0.0544 (0.0097)	0.0421 (0.0032)	θ_2	0.0542 (0.0037)
b_3	−0.3448 (0.1671)			β_{11}	0.0059 (0.0005)
A_{11}	−2.6164 (5.7308)	−2.6044 (3.4540)	−0.0755 (0.0224)	β_{22}	0.0068 (0.0005)
A_{21}	0.3357 (0.6079)	−2.6083 (3.4837)	3.9014 (0.3429)	σ_{31}	11.8613 (11.0421)
A_{31}	−2.7216 (5.2847)			λ_{11}	−19.3622 (11.0421)
A_{22}	−0.1972 (0.0932)	0.0679 (0.0198)	0.0276 (0.0189)	λ_{22}	−10.8795 (2.1794)
A_{32}	1.1258 (1.1833)			λ_{33}	−9.4415 (6.3615)
A_{33}	−0.2212 (0.0925)				
σ_{11}^2	6.2951[a] (8.3510)	0.2144[a] (0.4572)	0.0038[a] (0.0014)		
σ_{22}^2	0.1603[a] (0.2357)	2.6092[a] (0.7205)	3.8301[a] (1.1867)		
σ_{33}^2	0.0000[a] (0.0295)				
$\Phi_{0,1}$	0.1421 (0.0346)	0.1285 (0.2497)	0.0241 (0.0001)		
$\Phi_{0,2}$	0.0098 (0.0002)	−0.0033 (0.0093)	0.1994 (0.0154)		
$\Phi_{0,3}$	0.0624 (0.0197)				
$\Phi_{1,11}$	−10.1751 (8.7887)	−8.4277 (20.7884)	−1.6205 (0.0779)		
$\Phi_{1,21}$	−0.3934 (0.5276)	1.4735 (2.6585)	−15.1935 (1.0463)		
$\Phi_{1,31}$	−6.3263 (7.9648)				
$\Phi_{1,22}$	−0.0177 (0.0020)	−0.2077 (0.0907)	−0.2152 (0.0810)		
$\Phi_{1,32}$	0.1650 (0.4384)				
$\Phi_{1,33}$	−0.0884 (0.2462)				
κ		0.5826 (7.5578)	4.8265 (4.1876)		
θ		0.0327 (0.0301)	0.0195 (0.0091)		
σ		0.3969 (21.3490)	0.8931 (0.7155)		
λ_1		−0.3151 (28.9601)	1.0007 (0.0000)		
χ^2	47.504	90.031	45.190		242.160
df	17	24	24		30
z	5.231	9.531	3.058		27.390

[a] $\times 10^{-4}$. The table presents results of estimation of four term structure models utilizing the score generator indexed 11110000. The models are indexed as $A_m(n)Q(h)I(k)$, where $Q(h)$ denotes h quadratic factors, $A_m(n)$ denotes n affine factors of which m are square-root processes, and $I(k)$ denotes k inverted square root processes. The simulation size, N, is set to 50,000.

Results of specification tests are presented in Table 1. The table displays results for four models; $A_0(0)Q(3)I(0)$, $A_1(1)Q(2)I(0)$, $A_0(0)Q(2)I(1)$, and $A_2(3)Q(0)I(0)$. The models $A_0(0)Q(3)I(0)$ and $A_2(3)Q(0)I(0)$ are presented as benchmark cases; the results of Ahn et al. (2001) suggest that the quadratic model strongly outperforms the affine model in fitting the yield dynamics. Their results are confirmed here in a different data set covering an extended time period. As indicated by the z-statistic, the quadratic model provides a much better fit to the data than the affine model. The z-statistic for the quadratic model is 5.231, compared to 27.390 for the preferred affine model. These results suggest that the affine model has considerable difficulty in capturing term structure dynamics, even for a relatively simple SNP specification. This result supports the claim of Dai and Singleton (2000), who state that the tradeoff between specification of heteroskedastic volatility and correlation among the factors significantly impacts the ability of the affine model to fit term structure dynamics.

The second column presents results for a specification with two quadratic factors augmented by a square-root factor, $A_1(1)Q(2)I(0)$. The results indicate that, while this model improves upon the $A_2(3)Q(0)I(0)$ specification, it clearly sacrifices fit relative to the fully-specified quadratic model. The z-statistic for the model is 9.531, indicating that the quadratic model is strongly preferred on a degrees-of-freedom-adjusted basis. Since the $A_1(1)Q(2)I(0)$ and $A_0(0)Q(3)I(0)$ are similar in their mechanisms and abilities to generate heteroskedastic volatility, this result indicates that the correlation among the state variables plays an extremely important role in terms of model fit.

Results for the fourth model, $A_0(0)Q(2)I(1)$ are presented in the third column. The specification test indicates that this model improves upon the pure quadratic specification, with a z-statistic of 3.058. The key feature of this model is its specification of heteroskedastic volatility through the inverted square root factor. Our results indicate that this feature is quite important for conditional mean and volatility dynamics; the z-statistic declines substantially relative to the remaining three models. The result also generally supports the findings of Aït-Sahalia (1996a, b) and Ahn and Gao (1999), who suggest incorporating nonlinearity in the diffusion of the state variables to better describe the dynamics of bond yields. In comparing the results of this model to the $A_1(1)Q(2)I(0)$ model, it becomes apparent how potentially important this magnified level dependence is. Both models lack the correlation among the state variables, and this shortcoming severely impacts the $A_1(1)Q(2)I(0)$ model's fit. However, the $A_0(0)Q(2)I(1)$ model is able to improve upon the quadratic model despite sacrificing this correlation structure. This result suggests that flexibility in the specification of the level dependence of volatility may be more important in fitting the term structure than flexibility in the specification of correlation among the state variables.

To gain some better insight into the dimensions along which these models improve performance, we examine t-statistics for the scores of the model with respect to the SNP parameters. Consequently, we are able to analyze which parameters and thus which features of the data impair a given term structure model's ability to capture yield dynamics. These diagnostics are presented in Table 2. For the 11110000 score, there are two sets of parameters. The parameters $\psi(1) - \psi(12)$ represent the VAR terms in the score generator, and thus represent the conditional mean of the process. The $\tau(1) - \tau(30)$ terms are the GARCH parameters, which model the conditional volatility in the term structure.

The diagnostics in Table 2 confirm the conjecture that the incorporation of an inverted square root factor improves the fit of conditional volatility. Six of the scores with respect to the τ terms result in t-ratios greater than 2.0 for the $A_0(0)Q(3)I(0)$ model; in contrast, only two of the t-ratios for the $A_0(0)Q(2)I(1)$ model are greater than 2.0. Thus, the diagnostics suggest that while the hybrid quadratic-inverted square root model cannot completely capture the conditional dynamics of the yield curve, it appears to capture most of the relevant features of the data. The hybrid quadratic-square-root model performs much more poorly than the pure quadratic model, with thirteen of the

Table 2
EMM diagnostics of term structure models 11110000 score

Coefficient	$A_0(0)Q(3)I(0)$	$A_1(1)Q(2)I(0)$	$A_0(0)Q(2)I(1)$	$A_2(3)Q(0)I(0)$
$\psi(1)$	−1.807	−0.885	−0.830	−2.146
$\psi(2)$	1.786	1.053	0.431	3.145
$\psi(3)$	−0.673	−0.277	−0.536	−2.145
$\psi(4)$	1.540	0.958	0.095	0.864
$\psi(5)$	−1.744	−0.701	−0.030	−0.305
$\psi(6)$	0.585	0.257	0.012	0.033
$\psi(7)$	1.363	0.609	0.118	0.741
$\psi(8)$	−1.406	−0.415	−0.195	−0.145
$\psi(9)$	0.393	0.224	0.126	−0.067
$\psi(10)$	1.204	0.365	0.025	0.493
$\psi(11)$	−1.217	−0.347	−0.396	0.149
$\psi(12)$	0.249	0.299	0.272	−0.267
$\tau(1)$	1.973	1.418	0.898	3.897
$\tau(2)$	−0.808	−0.313	0.339	−6.067
$\tau(3)$	3.504	5.037	1.748	9.735
$\tau(4)$	1.906	−3.085	−0.035	−1.282
$\tau(5)$	−1.450	2.254	1.085	1.260
$\tau(6)$	0.596	0.699	−1.375	−2.111
$\tau(7)$	2.472	2.164	2.022	4.615
$\tau(8)$	−0.956	0.156	0.425	−7.802
$\tau(9)$	2.287	4.845	1.941	10.895
$\tau(10)$	2.226	−2.384	0.122	−0.538
$\tau(11)$	−1.486	1.584	0.927	1.781
$\tau(12)$	0.720	0.711	−0.631	−3.541
$\tau(13)$	1.771	1.848	1.749	3.459
$\tau(14)$	−0.490	−0.387	0.114	−5.933
$\tau(15)$	2.331	5.032	1.293	9.445
$\tau(16)$	1.408	−2.295	−0.148	−0.834
$\tau(17)$	−1.642	1.330	1.025	1.245
$\tau(18)$	1.122	0.824	−0.823	−1.846
$\tau(19)$	1.795	2.227	2.227	3.414
$\tau(20)$	0.168	−0.352	0.787	−6.696
$\tau(21)$	1.438	3.978	0.246	9.192
$\tau(22)$	1.552	−2.040	0.101	−0.812
$\tau(23)$	−1.720	1.250	0.776	1.034
$\tau(24)$	1.509	0.294	−0.647	−1.379
$\tau(25)$	1.740	1.520	1.273	3.918
$\tau(26)$	−1.189	−0.264	0.169	−7.845
$\tau(27)$	3.030	5.025	1.836	10.277
$\tau(28)$	1.907	−2.367	0.303	−0.809
$\tau(29)$	−1.495	1.516	0.894	1.475
$\tau(30)$	0.841	1.017	−0.935	−2.275

The table presents diagnostics for EMM scores evaluated for the score generating model 11110000. The coefficients labeled $\Psi(k)$ denotes the VAR terms of the SNP score generator and $\tau(k)$ denote the ARCH and GARCH terms of the SNP score generator. The table presents t-statistics for the test of the null hypothesis that the score with respect to the parameter $\Psi(k)$ or $\tau(k)$ is equal to 0.

Table 3
Specification tests of term structure models 11114300 score, N=50000

Coefficient	$A_0(0)Q(3)I(0)$	$A_1(1)Q(2)I(0)$	$A_0(0)Q(2)I(1)$	Coefficient	$A_2(3)Q(0)I(0)$
δ_0	0.0341 (0.0026)	0.0650 (0.0034)	0.0450 (0.0017)	κ_{11}	0.0052 (0.0018)
Ψ_{21}	−0.3230 (0.0925)	−7.9390 (0.0456)	−8.9953 (0.0354)	κ_{21}	−0.2037 (0.0949)
Ψ_{31}	0.2249 (0.1173)			κ_{22}	0.2442 (0.0157)
Ψ_{32}	−0.9787 (0.0071)			κ_{33}	5.1423 (0.4997)
b_1	−0.0501 (0.0088)	0.0564 (0.0004)	−0.0034 (0.0000)	θ_1	0.0088 (0.0007)
b_2	0.0890 (0.0005)	0.0583 (0.0006)	0.0395 (0.0004)	θ_2	0.0691 (0.0026)
b_3	−0.4872 (0.0022)			β_{11}	0.0038 (0.0001)
A_{11}	−2.4057 (0.1020)	−2.4308 (0.0027)	0.0725 (0.0011)	β_{22}	0.0054 (0.0003)
A_{21}	0.2932 (0.0225)	−2.5089 (0.0076)	4.0242 (0.0299)	σ_{31}	12.1908 (19.3312)
A_{31}	−2.2252 (0.0667)			λ_{11}	−19.1303 (6.1642)
A_{22}	−0.1829 (0.0011)	0.0701 (0.0014)	0.0352 (0.0018)	λ_{22}	−8.0897 (2.1924)
A_{32}	1.1056 (0.0074)			λ_{33}	−9.6907 (1.5573)
A_{33}	−0.2182 (0.0033)				
σ^2_{11}	7.9410[a] (0.8346)	0.2334[a] (0.0205)	0.0036[a] (0.0004)		
σ^2_{22}	0.1702[a] (0.0139)	2.5390[a] (0.0597)	3.9319[a] (0.0672)		
σ^2_{33}	0.0000[a] (0.0042)				
$\Phi_{0,1}$	0.1256 (0.0010)	0.1545 (0.0025)	0.0325 (0.0000)		
$\Phi_{0,2}$	0.0129 (0.0000)	−0.0033 (0.0005)	0.1811 (0.0005)		
$\Phi_{0,3}$	0.0203 (0.0012)				
$\Phi_{1,11}$	−10.3223 (0.0423)	−8.9150 (0.1349)	−1.6614 (0.0023)		
$\Phi_{1,21}$	−0.3520 (0.0003)	−1.3350 (0.0152)	−14.3815 (0.0350)		
$\Phi_{1,31}$	−5.7044 (0.0292)				
$\Phi_{1,22}$	−0.0188 (0.0001)	−0.2066 (0.0023)	−0.2062 (0.0027)		
$\Phi_{1,32}$	0.1761 (0.0015)				
$\Phi_{1,33}$	−0.0839 (0.0017)				
κ		0.5826 (0.3686)	4.1039 (0.3268)		
θ		0.0327 (0.0026)	0.0246 (0.0010)		
σ		0.3969 (0.9196)	0.7875 (0.0483)		
λ_1		−0.3151 (1.1347)	1.5458 (0.4113)		
χ^2	100.610	161.026	129.411		306.863
df	29	36	36		42
z	9.403	14.734	11.009		28.899

[a] $\times 10^{-4}$. The table presents results of estimation of four term structure models utilizing the score generator indexed 11114300. The models are indexed as $A_m(n)Q(h)I(k)$, where $Q(h)$ denotes h quadratic factors, $A_m(n)$ denotes n affine factors of which m are square-root processes, and $I(k)$ denotes k inverted square root processes. The simulation size, N, is set to 50,000.

τ score t-ratios exceeding 2.0 in magnitude. The $A_2(3)Q(0)I(0)$ model has the greatest difficulty capturing conditional volatility, exhibiting eighteen t-ratios that exceed 2.0 in absolute value.

In summary, our specification tests indicate that modeling term structure dynamics using a combination of different types of factors may allow us to better capture yield dynamics. In particular, utilizing quadratic and inverted square-root factors improves substantially upon the ability of quadratic factors alone to fit the conditional dynamics of the yield curve. Much of this improvement is through the hybrid model's ability to produce conditional volatility that is consistent with the data. However, these results hinge on ignoring deviations from conditional normality in the data. In the next section, we examine the impact of incorporating this feature through augmenting our score generator with a Hermite polynomial.

4.1.2. Results with the 11114300 score

Estimation results with the 11114300 score generator, which more fully characterizes the data, are presented in Table 3. As expected, all four models estimated experience a deterioration in goodness of fit as a result of the incorporation of the Hermite polynomial terms. As in the previous section, the quadratic model, $A_0(0)Q(3)I(0)$ continues to fit the data better than the affine model, $A_2(3)Q(0)I(0)$. However, the measure of fit falls sharply; the z-statistic of the quadratic model is 9.4, compared to 5.3 using the 11110000 score. Consequently, the addition of Hermite polynomial terms appear to severely impact the model's ability to capture conditional yield dynamics.

However, the $A_0(0)Q(2)I(1)$ model suffers an even more dramatic deterioration in fit. The model's z-statistic rises from 3.1 to 11.0; the deviations from normality significantly impact the model's ability to fit the data. In contrast, the $A_1(1)Q(2)I(0)$ model does not experience as drastic a reduction in fit; the model's z-statistic rises from 9.5 to 14.7. The rise in the z-statistic for the $A_1(1)Q(2)I(0)$ model roughly parallels that of the of the $A_0(0)Q(3)I(0)$ model. This result suggests that the addition of the Hermite polynomial terms do not impact the fit of this hybrid model much relative to the quadratic model. The sharp drop in fit for the hybrid quadratic-enverted square-root model in conjunction with the milder drop in fit of the quadratic and quadratic-square-root models, suggest that flexibility in modeling correlation among the state variables is important for capturing the shape characteristics of the conditional density.[16]

We analyze the t-ratios for the scores of the 11114300 to help us assess the validity of our conjectures. These results are presented in Table 4, and confirm the conclusion that the quadratic model is best able to capture deviations from conditional normality. None of the scores with respect to the Hermite terms $A(2) - A(13)$ are statistically significant when the quadratic model is estimated. In contrast, the $A_0(1)Q(2)I(0)$ and $A_0(0)Q(2)I(1)$ models produce three and two significant Hermite term scores respectively. Thus, the indication is that the quadratic model is better able to capture the shape features of the conditional density.

However, the improvement in fitting the Hermite scores appears to come at a large cost in terms of fitting the conditional mean. The quadratic model violates four of the conditional mean terms, whereas each of the hybrid models violate one of these terms. Thus, the t-ratios suggest that the hybrid models maintain somewhat better fit of the conditional mean dynamics. However, in the presence of deviations from conditional normality, the hybrid models sacrifice a considerable degree of conditional volatility fit. Whereas only two of the t-ratios with respect to the τ terms exceed 2.0 for the quadratic model, the $A_0(1)Q(2)I(0)$ and $A_0(0)Q(2)I(1)$ produce nine and five significant t-ratios

[16] We repeat the analyses of the 11114300 score with a simulation length $T = 100000$ in untabulated results. The qualitative conclusion of the specification tests is unchanged. The model $A_0(0)Q(3)I(0)$ produces the lowest z-statistic, followed by the $A_0(0)Q(2)I(1)$. The statistics for these models rise slightly, suggesting that the overall fit is slightly worse. However, the differences do not appear to be material.

Table 4
EMM diagnostics of term structure models: 11114300 score

Coefficient	$A_0(0)Q(3)I(0)$	$A_1(1)Q(2)I(0)$	$A_0(0)Q(2)I(1)$	$A_2(3)Q(0)I(0)$
$A(2)$	0.134	0.465	−0.215	−2.416
$A(3)$	−0.689	1.497	−1.024	1.061
$A(4)$	−0.478	0.261	0.806	−1.622
$A(5)$	−1.764	3.464	−1.482	0.021
$A(6)$	0.728	5.463	2.354	7.105
$A(7)$	1.212	2.316	1.431	4.825
$A(8)$	−0.353	−0.893	−0.628	−1.739
$A(9)$	0.628	1.643	0.073	1.759
$A(10)$	−0.117	−0.793	−1.468	−0.253
$A(11)$	1.124	2.365	0.428	2.192
$A(12)$	1.848	4.538	2.384	5.476
$A(13)$	1.307	1.852	1.777	3.325
$\psi(1)$	−0.531	1.647	2.443	−1.888
$\psi(2)$	0.780	−0.779	−1.886	1.342
$\psi(3)$	0.838	1.383	1.408	−0.612
$\psi(4)$	2.123	0.772	−0.031	1.109
$\psi(5)$	−1.713	−0.511	0.079	−0.565
$\psi(6)$	0.599	0.138	0.185	−0.200
$\psi(7)$	2.399	0.453	0.269	1.060
$\psi(8)$	−1.867	−0.269	−0.182	−0.767
$\psi(9)$	0.442	0.164	0.328	−0.194
$\psi(10)$	2.518	0.230	0.256	0.921
$\psi(11)$	−2.000	−0.237	−0.442	−0.854
$\psi(12)$	0.307	0.191	0.471	−0.227
$\tau(1)$	2.247	2.055	−0.660	4.552
$\tau(2)$	−0.297	−0.743	−1.483	−5.724
$\tau(3)$	1.250	4.517	2.706	9.906
$\tau(4)$	0.264	−3.490	−1.763	−2.550
$\tau(5)$	−0.413	0.979	1.246	2.096
$\tau(6)$	−0.232	1.118	−1.587	−1.343
$\tau(7)$	2.435	2.992	1.369	5.603
$\tau(8)$	−0.481	−0.304	−1.526	−5.903
$\tau(9)$	0.775	4.427	3.725	11.060
$\tau(10)$	0.605	−2.781	−1.719	−2.061
$\tau(11)$	−0.644	1.029	1.606	2.725
$\tau(12)$	−0.160	1.085	−1.306	−2.110
$\tau(13)$	1.337	2.177	0.422	3.714
$\tau(14)$	−0.388	−0.686	−1.632	−4.361
$\tau(15)$	0.528	4.623	2.993	8.395
$\tau(16)$	−0.065	−2.833	−1.696	−1.792
$\tau(17)$	−0.252	1.166	1.759	2.169
$\tau(18)$	−0.240	0.983	−1.651	−1.448
$\tau(19)$	1.037	2.341	0.636	3.495
$\tau(20)$	−0.322	−0.688	−1.691	−5.752
$\tau(21)$	0.033	4.397	2.060	9.214
$\tau(22)$	0.124	−2.638	−1.295	−1.763
$\tau(23)$	−0.520	1.523	1.865	2.378
$\tau(24)$	0.140	0.143	−1.801	−1.845
$\tau(25)$	2.285	2.516	0.438	5.232
$\tau(26)$	−0.370	−0.570	−1.597	−6.137
$\tau(27)$	0.916	4.557	3.143	10.265
$\tau(28)$	0.693	−2.436	−1.422	−2.141
$\tau(29)$	−0.593	0.792	1.307	2.493
$\tau(30)$	−0.128	1.230	−1.427	−2.161

The table presents diagnostics for EMM scores evaluated for the score generating model 11114300. The coefficients labeled $A(k)$ denote the coefficients of the Hermite polynomial of the SNP score generator, $\Psi(k)$ denotes the VAR terms of the SNP score generator and $\tau(k)$ denote the ARCH and GARCH terms of the SNP score generator. The table presents t-statistics for the test of the null hypothesis that the score with respect to the parameter $A(k)$, $\Psi(k)$, or $\tau(k)$ is equal to 0.

respectively. Thus, the quadratic model appears to have a substantial edge in fitting the higher conditional moments.

These diagnostics indicate that incorporating the Hermite polynomial terms can have a significant impact on which models are able to better describe the data. In particular, the quadratic model and hybrid quadratic-inverted square root model reverse rankings in terms of fit of the more complicated score. Interestingly, the edge given to the quadratic model seems to come at a large cost in terms of fitting the conditional mean of the density. The model apparently attempts to fit conditional volatility and conditional non-normality more aggressively than the conditional mean. However, the apparent overall conclusion is that while the inverted square root factor may better capture conditional volatility in the absence of deviations from normality, correlation among the state variables is important in modeling shape departures from a conditionally normal distribution.

4.2. Reprojection

Gallant and Tauchen (1998) present an additional diagnostic for gauging the performance of alternative structural models, termed reprojection. A detailed discussion of the method is provided in their paper. Conceptually, the method compares the conditional density for discretely sampled data that is implied by the structural model to a conditional density computed directly from the data. Closed-form solutions are not in general available for the conditional density implied by a structural model. However, we can set the structural parameters ρ to their EMM estimates $\hat{\rho}_n$, generate a large simulation, and fit an SNP model to the simulation. Gallant and Long (1997) prove, under regularity conditions, that the SNP density thus computed converges to the conditional density implied by the structural model.

Of immediate interest in eliciting the dynamics of observables are the first two one-step-ahead conditional moments

$$\mathscr{E}(y_0|y_{-L},\ldots,y_{-1}) = \int y_0 f_K(y_0|x_{-1},\hat{\theta})\,\mathrm{d}y_0$$

and

$$\mathrm{Var}(y_0|y_{-L},\ldots,y_{-1}) = \int [y_0 - \mathscr{E}(y_0|x_{-1})][y_0 - \mathscr{E}(y_0|x_{-1})]' f_K(y_0|x_{-1},\hat{\theta})\,\mathrm{d}y_0$$

where $x_{-1} = (y_{-L},\ldots,y_{-1})$, f_K represents the SNP auxiliary model, and $\hat{\theta}$ represents the SNP parameter estimates computed from the simulation. Thus the reprojection method provides some further insight into the performance of the models in capturing the conditional means, volatilities, and deviations from normality implied by the data.

Plots of the first two conditional (cross) moments implied by three of the models are presented in Figs. 2–10.[17] The figures depict the conditional moments implied by

[17] We do not provide reprojection results for the $A_1(1)Q(2)I(0)$ model since it is dominated by the $A_0(0)Q(2)I(1)$ model.

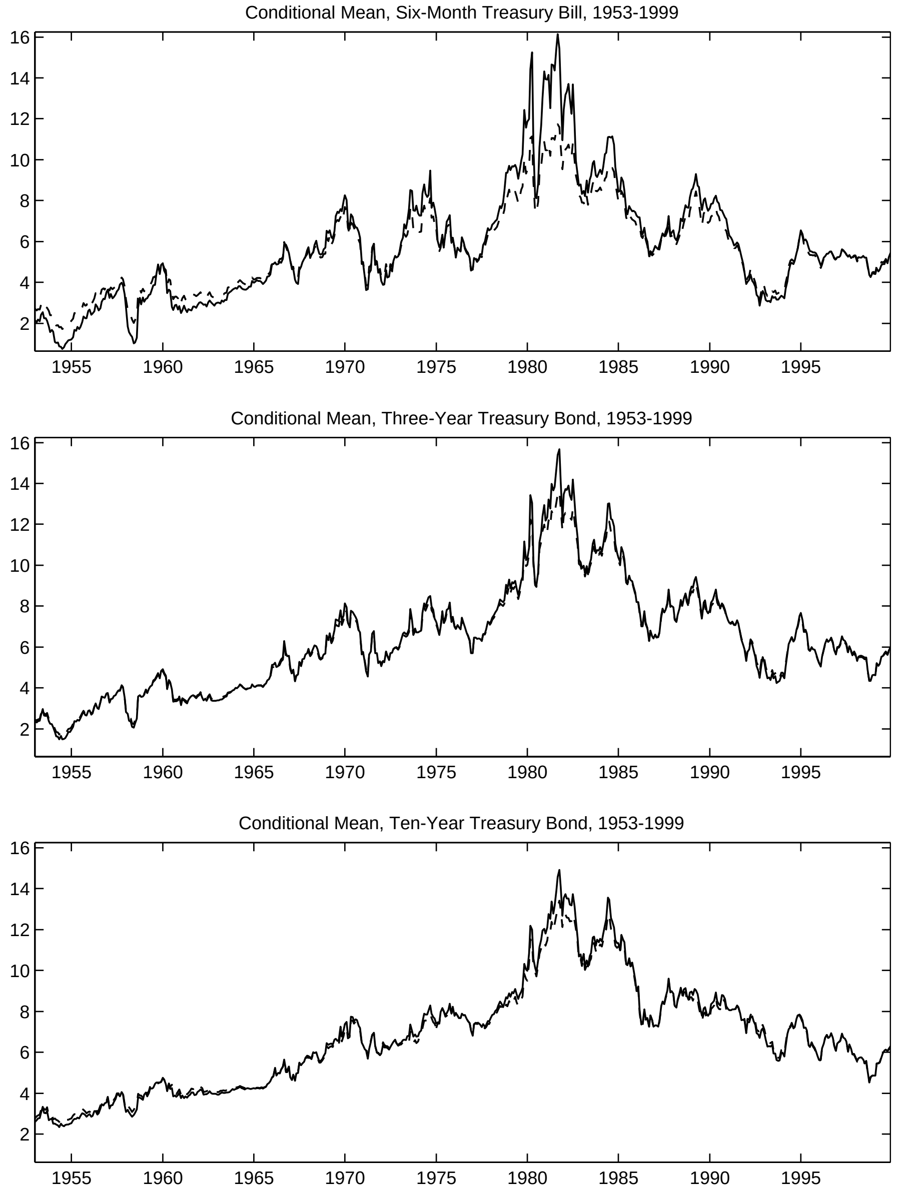

Fig. 2. Reprojected conditional mean: $A_0(0)Q(3)I(0)$. The plots present the reprojected conditional mean for the $A_0(0)Q(3)I(0)$ model against the projected conditional mean. The reprojected data are represented by dashed lines, whereas the projected data are represented by solid line.

the models (dashed lines) and the conditional moments implied by the score generator (solid lines). Figs. 2 through 4 present results for the fully specified three-factor quadratic model. In general, these graphs suggest that the model provides a reasonably

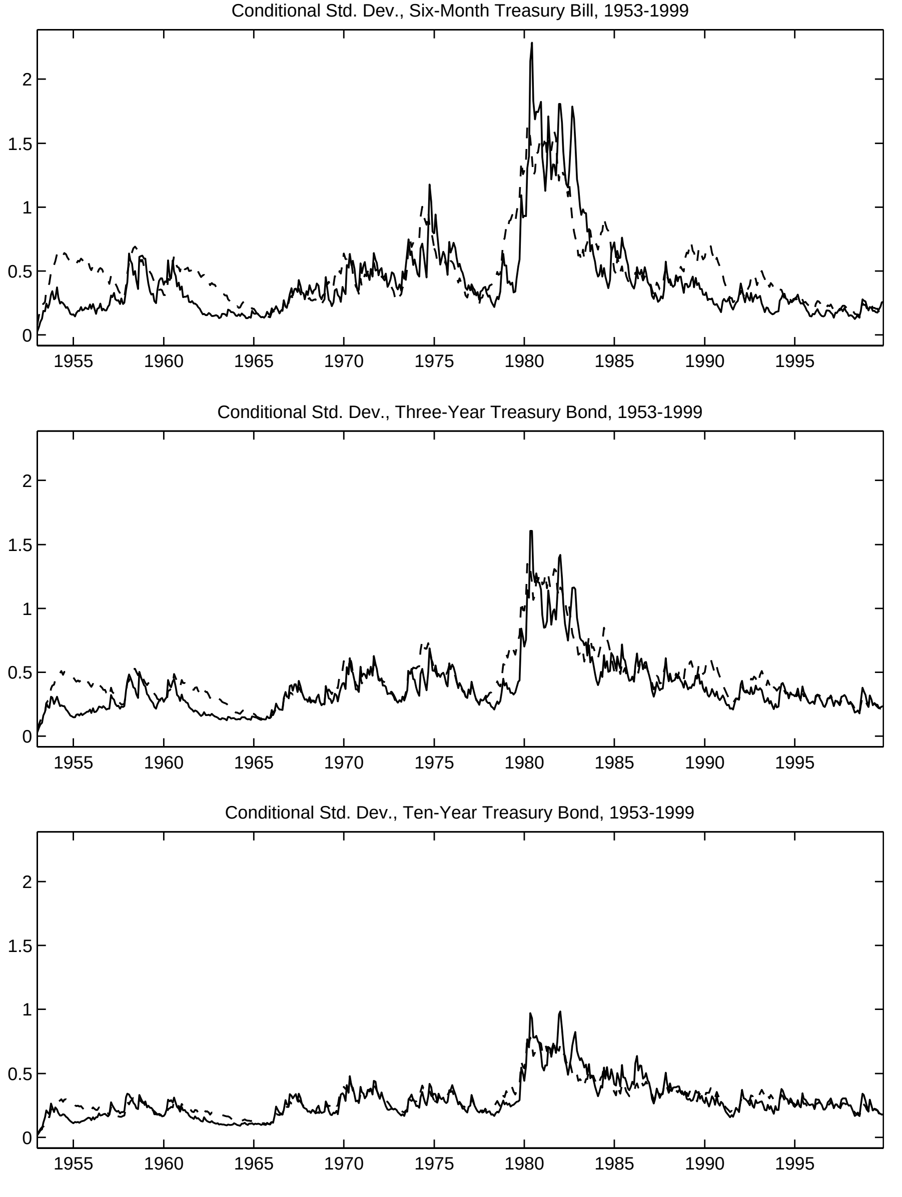

Fig. 3. Reprojected conditional volatility: $A_0(0)Q(3)I(0)$. The plots present the reprojected conditional volatility for the $A_0(0)Q(3)I(0)$ model against the projected conditional volatility. The reprojected data are represented by dashed lines, whereas the projected data are represented by solid line.

accurate description of the data. The model has some difficulty reproducing the high conditional means of the short- and long-term bonds in the high interest rate period of the early 1980s, but tracks conditional volatility quite well. Further, the model is able to generate the general shape of conditional volatility as well. The conditional

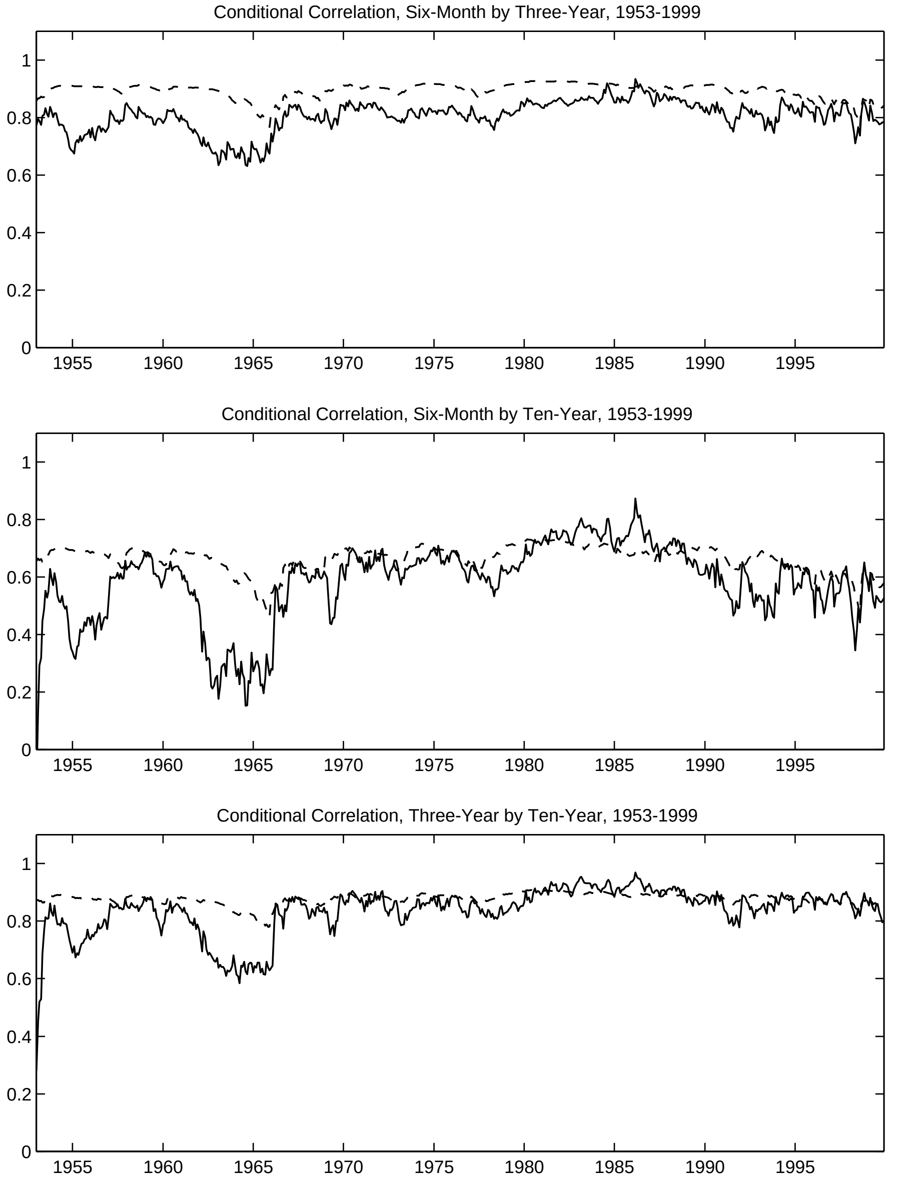

Fig. 4. Reprojected conditional correlation: $A_0(0)Q(3)I(0)$. The plots present the reprojected conditional correlation for the $A_0(0)Q(3)I(0)$ model against the projected conditional correlation. The reprojected data are represented by dashed lines, whereas the projected data are represented by solid line.

correlation implied by the model is smoother than the projected conditional correlations, particularly for the correlation of the short-term bond with the long-term bond. In the late 1960s, the model appears to oversmooth and overestimate correlations in particular.

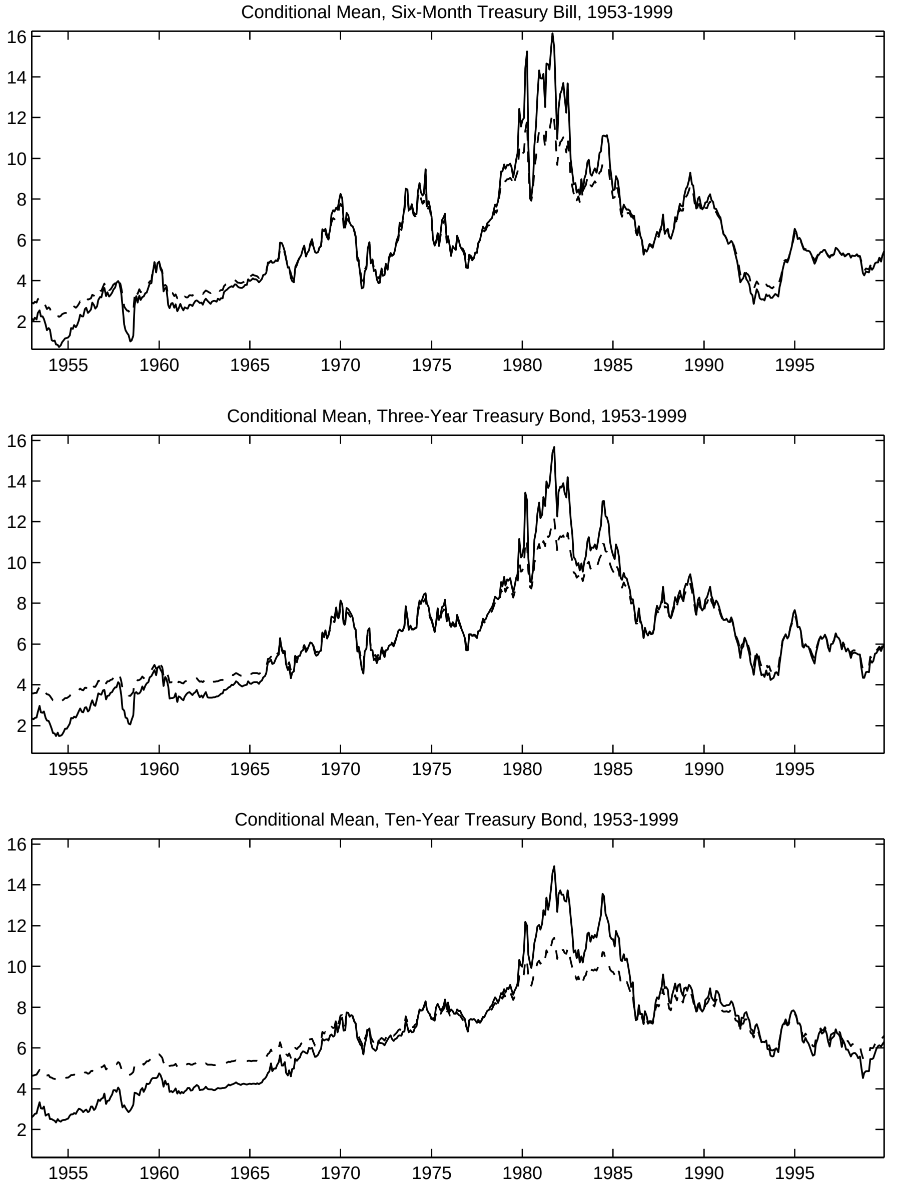

Fig. 5. Reprojected conditional mean: $A_0(0)Q(2)I(1)$. The plots present the reprojected conditional mean for the $A_0(0)Q(2)I(1)$ model against the projected conditional mean. The reprojected data are represented by dashed lines, whereas the projected data are represented by solid line.

Figs. 5 through 7 present plots of the reprojected conditional moments for the $A_0(0)Q(2)I(1)$ model. The reprojections indicate that the model appears to slightly underperform the quadratic model in terms of capturing the conditional mean of the

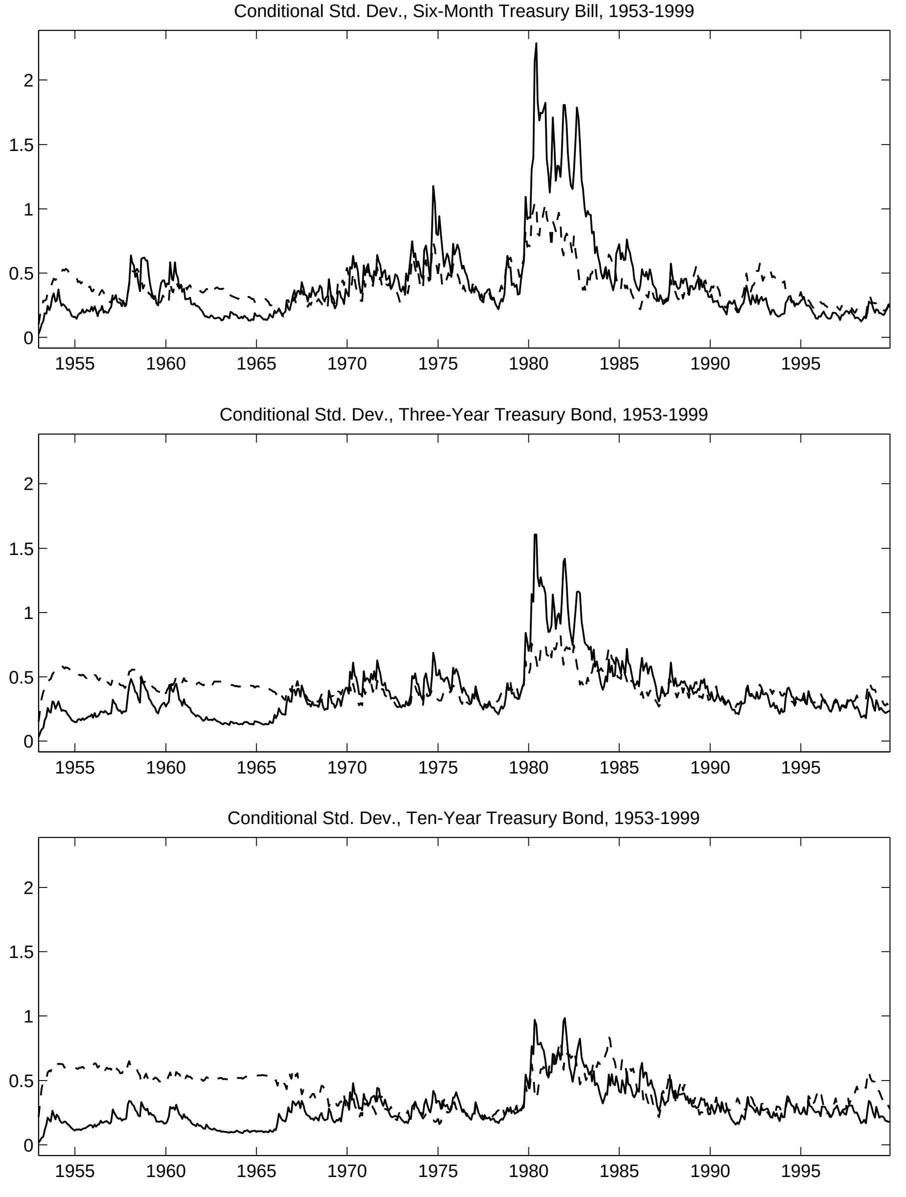

Fig. 6. Reprojected conditional volatility: $A_0(0)Q(2)I(1)$. The plots present the reprojected conditional volatility for the $A_0(0)Q(2)I(1)$ model against the projected conditional volatility. The reprojected data are represented by dashed lines, whereas the projected data are represented by solid line.

data. In particular, the model seems to overestimate conditional means in the early part of the sample period. The model has a similar problem with conditional volatility early in the sample period and cannot quite generate the level of volatility observed in the

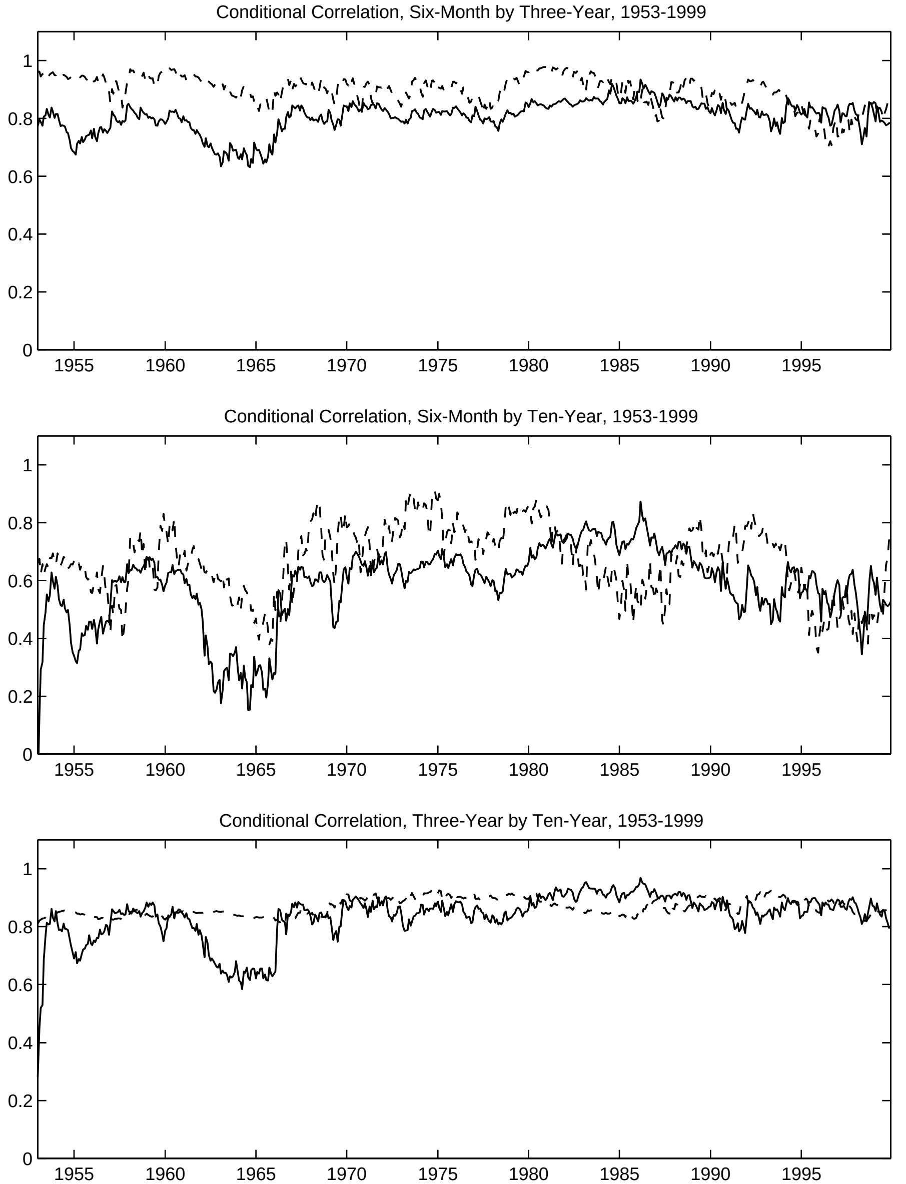

Fig. 7. Reprojected conditional correlation: $A_0(0)Q(2)I(1)$. The plots present the reprojected conditional correlation for the $A_0(0)Q(2)I(1)$ model against the projected conditional correlation. The reprojected data are represented by dashed lines, whereas the projected data are represented by solid line.

high-yield period of the early 1980s. However, the model does a somewhat better job than the pure quadratic model at capturing the conditional correlations in bond yields in the late 1960s; the overall pattern for the 6-month, 10-year correlation in particular seems better to represent conditional correlations in this period.

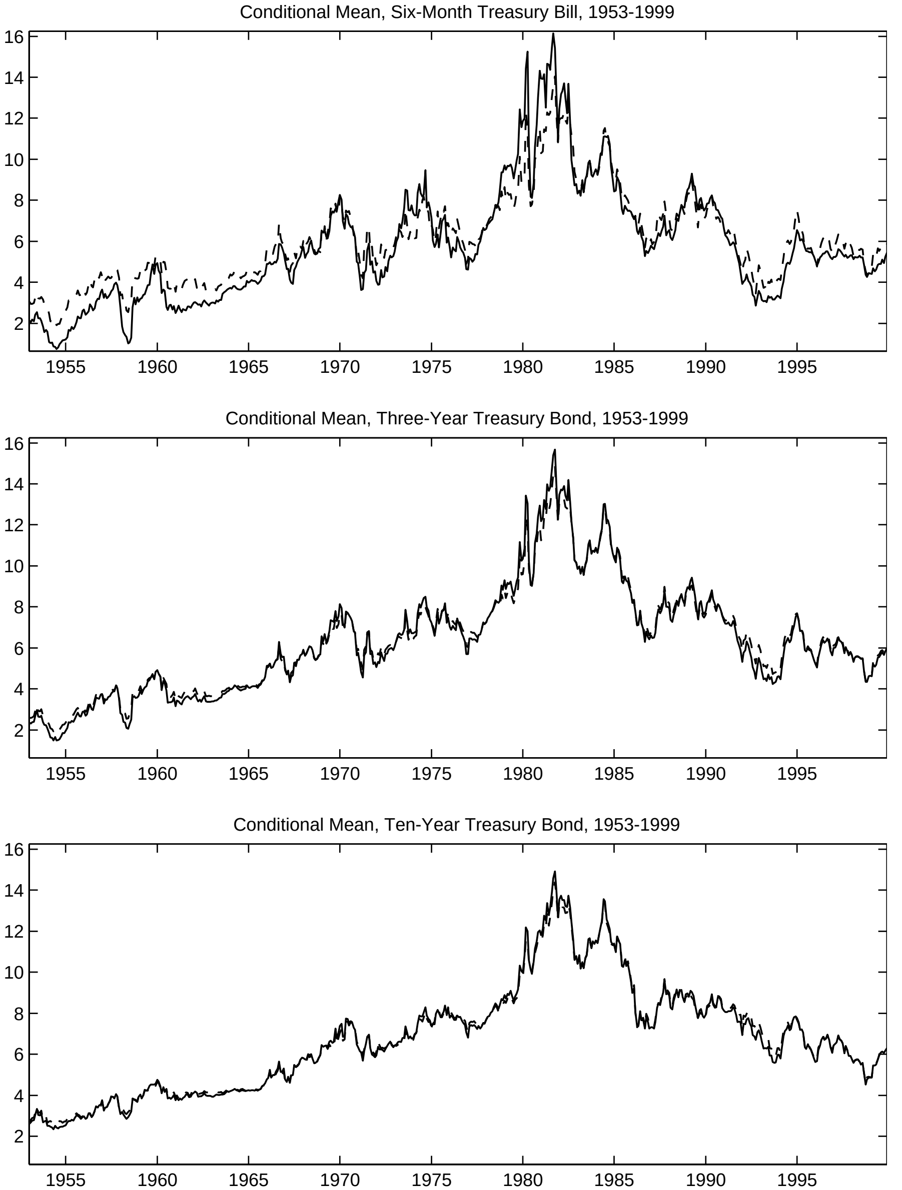

Fig. 8. Reprojected conditional mean: $A_2(3)Q(0)I(0)$. The plots present the reprojected conditional mean for the $A_2(3)Q(0)I(0)$ model against the projected conditional mean. The reprojected data are represented by dashed lines, whereas the projected data are represented by solid line.

Finally, Figs. 8 through 10 suggest that the fully specified affine model performs weakly in capturing conditional volatility and conditional correlation. The reprojected conditional mean plot suggests that the affine model is able to reproduce term structure mean dynamics fairly well. However, the model has considerably more difficulty re-

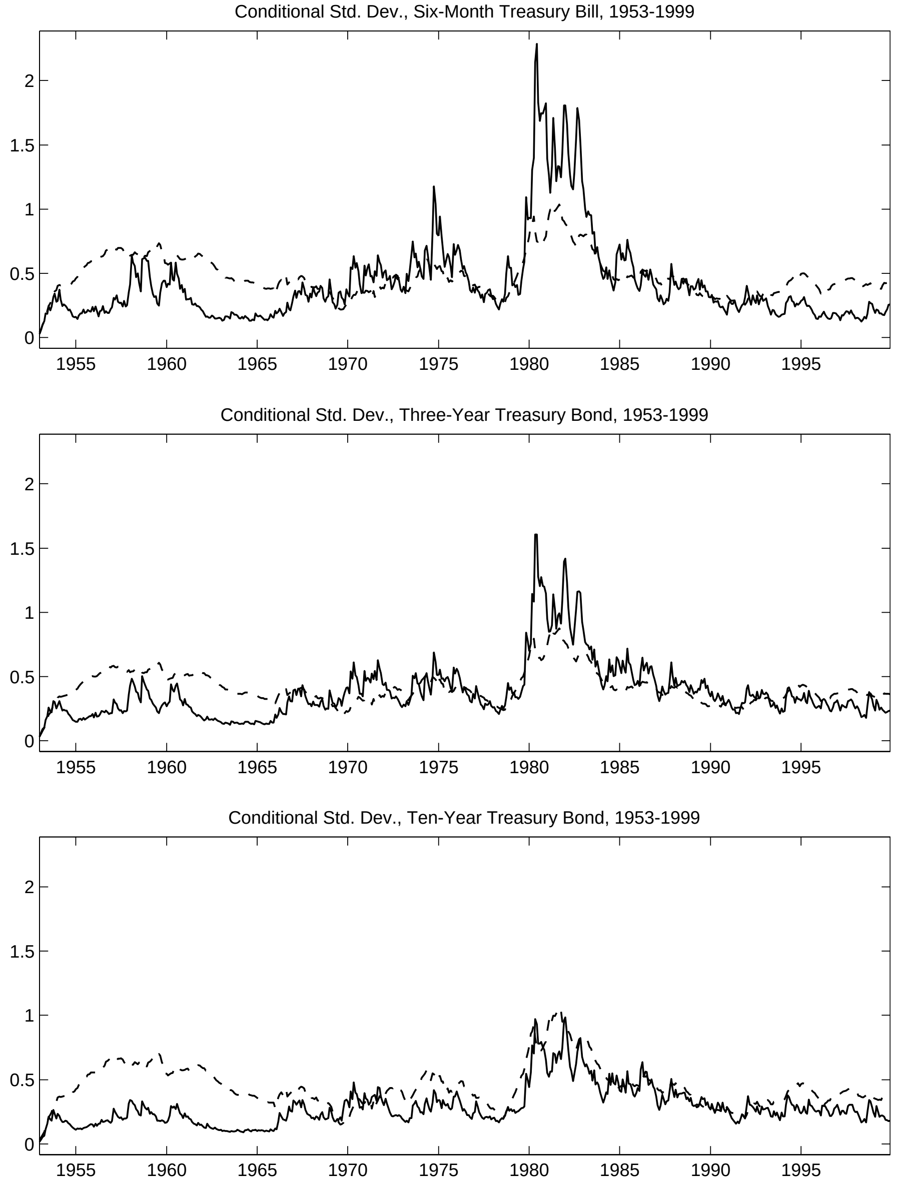

Fig. 9. Reprojected conditional volatility: $A_2(3)Q(0)I(0)$. The plots present the reprojected conditional volatility for the $A_2(3)Q(0)I(0)$ model against the projected conditional volatility. The reprojected data are represented by dashed lines, whereas the projected data are represented by solid line.

producing the conditional volatility of yields, and generates particularly smooth conditional correlations. Thus, as discussed previously, the reprojection results suggest that the weakness of the affine specification is in its ability to capture conditional second moments.

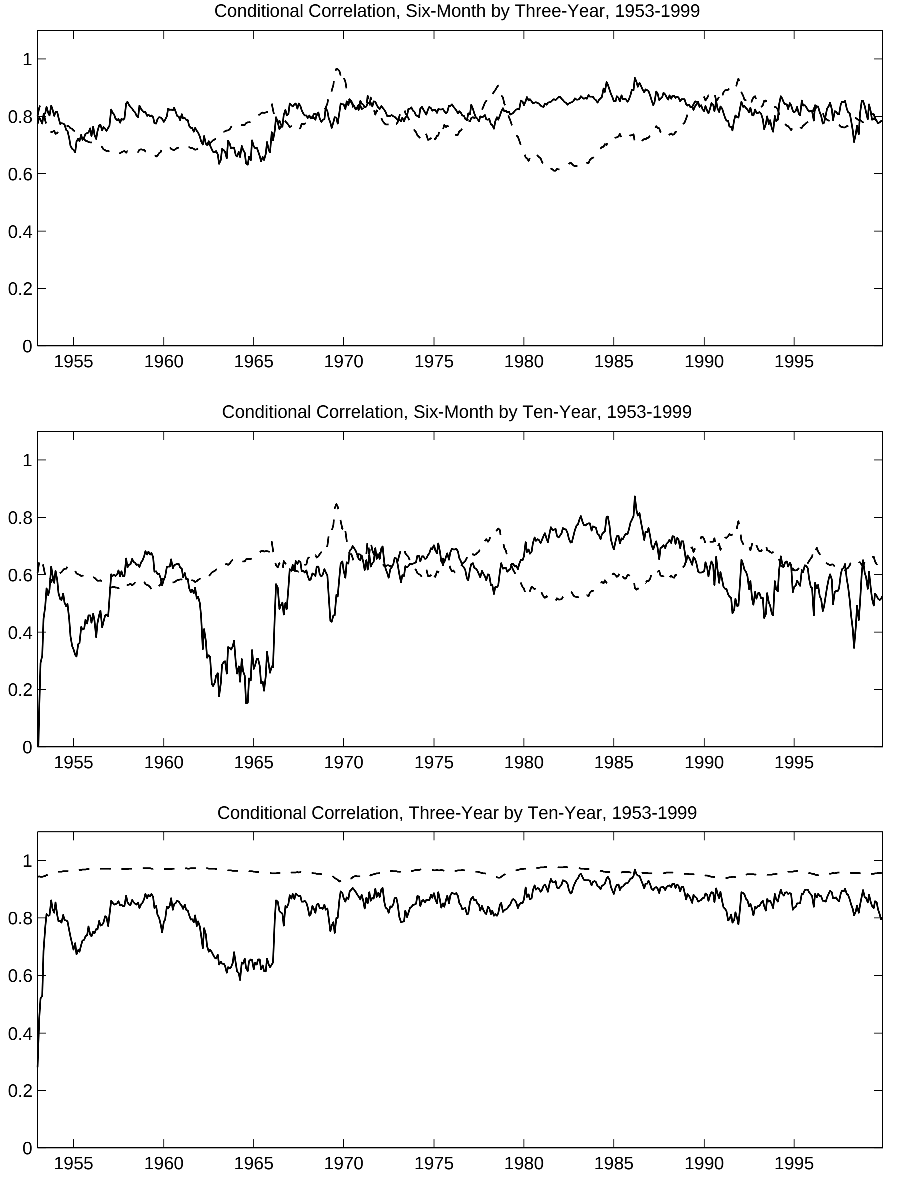

Fig. 10. Reprojected conditional correlation: $A_2(3)Q(0)I(0)$. The plots present the reprojected conditional correlation for the $A_2(3)Q(0)I(0)$ model against the projected conditional correlation. The reprojected data are represented by dashed lines, whereas the projected data are represented by solid line.

In summary, the results of the reprojection analysis conform largely to the results of the specification tests. The quadratic model performs quite well in capturing conditional second moments in this setting. The probable source of this performance is its maximal flexibility in specifying correlations among the state variables. The hybrid quadratic-inverted square-root model has more difficulty in capturing the conditional moments when deviations from conditional normality are present in the data.

5. Conclusion

This study confirms, outside of the affine factor specification, the claim made by Dai and Singleton (2000) that a trade-off exists between the factor level dependence of conditional variance in interest rates and the admissible structure of the factor correlation matrix. In our specification tests, we find that a hybrid quadratic-inverted-square-root model outperforms the pure quadratic model when the data are described by a conditionally normal GARCH(1,1) model. This result is achieved despite the fact that a more restrictive correlation structure among the state variables is forced by the hybrid model. The performance enhancement derives from stronger level dependence in the volatility of interest rates induced by the inverted square-root factor.

When we allow for deviations from conditional normality, the results reverse, and the quadratic model better describes the data than the quadratic-inverted-square-root model. In this setting, the improvement from a more flexible correlation specification dominates the improvements realized from inducing greater level dependence. In conjunction with the results with a less complicated conditional density model, these results suggest that care must be taken in specifying both level dependence and correlation structure. Both features should apparently be given similar weight in specifying factor dynamics for the purpose of modeling the term structure. We conjecture that a model that allows for a fully-specified correlation structure and increased level dependence would further improve our ability to fit the dynamics of the term structure. However, under current modeling techniques, such a model would have to rely on a numerical solution to a PDE for bond prices.

Acknowledgements

We are grateful to Mike Chernov, Eunho Chung, Jennifer Conrad, Qiang Dai, Lars Hansen, Philsang Lee, and Jinbum Choi for their useful discussions. We have also benefitted from the comments of seminar participants at the Indiana University Lunch Workshop and the Conference on Risk Neutral and Objective Probability Distributions, Duke University October 20–21, 2000. Comments from two anonymous referees are greatly appreciated. The usual disclaimer applies. The first author greatly acknowledges financial support from Korea University and the Com2Mac Center at POSTECH. This material is based upon work supported by the National Science Foundation under Grant No. 0000176.

References

Ahn, D.-H., Gao, B., 1999. A parametric non-linear model of term structure dynamics. Review of Financial Studies 12, 721–762.

Ahn, D.-H., Dittmar, R. F., Gallant, A.R., 2001. Quadratic term structure models: Theory and evidence. Review of Financial Studies 15, 243–288.

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

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

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

Balduzzi, P., Das, S., Foresi, S., Sundaram, R., 1996. A simple approach to three factor affine models of the term structure. Journal of Fixed Income 6, 43–53.

Beaglehole, D., Tenney, M., 1991. General solutions of some interest rate-contingent claim pricing equations. Journal of Fixed Income 1, 69–83.

Beaglehole, D., Tenney, M., 1992. A nonlinear equilibrium model of the term structure of interest rates: corrections and additions. Journal of Financial Economics 32, 345–454.

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

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

Chen, L., 1996. Stochastic Mean and Stochastic Volatility—A Three-Factor Model of the Term Structure of Interest Rates and Its Application to the Pricing of Interest Rate Derivatives. Blackwell, Oxford, UK.

Chen, R.R., Scott, L., 1992. Pricing interest options in a two-factor Cox-Ingersoll-Ross model of the term structure. Review of Financial Studies 5, 613–636.

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

Cox, J.C., Ingersoll, J.E., Ross, S.A., 1985. A theory of the term structure of interest rates. Econometrica 53, 385–406.

Dai, Q., Singleton, K., 2000. Specification analysis of affine term structure models. Journal of Finance 55, 1943–1978.

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

Engle, R., Ng, V., 1993. Time-varying volatility and the dynamic behavior of the term structure. Journal of Money, Credit, and Banking 25, 336–349.

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

Gallant, A.R., Hsu, C.-T., Tauchen, G., 1999. Using daily range data to calibrate volatility diffusions and extract the forward integrated variance. The Review of Economics and Statistics 81 (4), 617–631.

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

Gallant, A.R., Tauchen, G., 1998. Reprojecting partially observed systems with application to interest rate diffusions. Journal of the American Statistical Association 93, 10–24.

Gallant, A.R., Tauchen, G., 2001. Efficient method of moments. Duke University, unpublished manuscript.

Jegadeesh, N., Pennachi, G., 1996. The behavior of interest rates implied by the term structure of Eurodollar futures. Journal of Financial and Quantitative Analysis 26, 426–446.

Langetieg, T., 1980. A multivariate model of the term structure. Journal of Finance 35, 71–97.

Leippold, M., Wu, L., 2000. Quadratic term structure models. Fordham University, unpublished manuscript.

Longstaff, F., 1989. A nonlinear general equilibrium model of the term structure of interest rates. Journal of Finance 23, 1259–1282.

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

McCulloch, J., Kwon, H., 1993. U.S. term structure data, 1947–1991. Ohio State University, unpublished manuscript.

Merton, R., 1973. An intertemporal capital asset pricing model. Econometrica 41, 867–887.

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

Sun, T., 1992. Real and nominal interest rates: A discrete-time model and its continuous time limit. Review of Financial Studies 5, 581–612.

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