Purebred or Hybrid?:
Reproducing the Volatility in Term Structure Dynamics

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

First Draft: September, 2000

Abstract

1 Introduction

Understanding the dynamics of the term structure of interest rates is important to researchers as well as practitioners because of these dynamics' implications for macroeconomic policy and microeconomic decision making of economic agents in an economy. The new millenium has already predicted and experienced turbulence in fixed-income security markets in response to the 'soft landing' of the US economy and anticipated inflation induced by soaring oil prices. Consequently, significant attention has been paid in both the academic and practitioner literature to attempting to describe term structure dynamics.

The empirical evidence suggests that the uncertainty or volatility in interest rate dynamics is time-varying and probably stochastic [e.g. Longstaff and Schwartz (1992), Engle and Ng (1993), Gallant and Tauchen (1996)]. This stream of literature relies upon the premise that i) there exists strong time-variation in term premia and ii) these time-varying premia are initially induced by time-varying volatility of interest rates and/or associated state variables which govern the stochastic dynamics of the term structure. Despite the major progress made in documenting the empirical dynamics of the second moment of Treasury securities, the progress in connecting the observed empirical dynamics of the second moment and existing parametric term structure models is relatively limited. Even though many studies investigate the empirical ability of alternative specifications of the theoretical stochastic differential equation of *state variables* to explain the time-varying volatility of the short rate, there are only a few studies that investigate the empirical performance of parametric term structure models in explaining the time-varying volatilities of the yield-to-maturities.[1] The latter set of studies concerns the empirical performance of the models in both the cross-sectional and time series dimensions, or equivalently the models' performance under the equivalent martingale measure. Thus, these studies may be viewed as a more rigorous empirical investigation of the underlying models. Exceptions are the recent studies of Dai and Singleton (2000) and Ahn, Dittmar and Gallant (2000, ADG hereafter). Dai and Singleton (2000) 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 1980's through the 1990's. ADG document that Quadratic Term Structure Models (*QTSM*s henceforth) outperform *ATSM*s empirically, but that neither the *ATSM*s nor the *QTSM*s can explain the stochastic features of the volatilities of the yields, in particular the short rate.

To illustrate this point, Figure 1 depicts the term structure of volatility, e.g., the locus of the volatility of bond yields as a function of time-to-maturity. We observe that *ATSM*s produce lower levels of volatility across all maturities, and in addition, have difficulty in recovering the shape of the term structure of volatility. *QTSM*s improve upon the *ATSM*s performance in both dimensions, but

---

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

still suffer from the inability to reproduce the large magnitude of volatility across all maturities. In particular, the $QTSM$ is relatively successful in explaining the observed shape of the term structure of volatility dynamics, but it fails significantly to generate excessive volatility levels in the interest rates over different horizons. In summary, we suggest that the representative parametric term structure models have shortfalls in fitting the observed volatilities of the yields-to-maturity across time to maturity.

To understand the shortcomings of extant parametric models in their specifications of volatility, we need to investigate existing term structure models and their relative strengths and weaknesses from the viewpoint of their implications for term structure volatility. More than two decades of extensive studies in term structure dynamics can be classified into three alternative classes.[2] First, the most popular class of term structure models is the class of $ATSM$s, which designate an affine association between the underlying state variables and yields. This class includes the seminal works of Vasicek (1977) and Cox, Ingersoll and Ross (1985, CIR hereafter).[3] Duffie and Kan (1996) make a significant contribution to the literature by generalizing $ATSM$s, and summarize the primitive assumptions underlying this class of models. Dai and Singleton (2000) characterize the admissibility of $ATSM$s and explore a maximally flexible $ATSM$ that empirically nests all other existing $ATSM$s as a sub-family. Dai and Singleton also pioneer a new focus in the empirical literature by implementing the $ATSM$s based on theoretical analysis of the minimal conditions necessary for for identifying the $ATSM$s. Even though their analysis exhibits relatively promising empirical performance of this class for late 1980's and early 1990's, Gallant and Tauchen (1997), and ADG (2000) reject all versions of $ATSM$s using data including the 1970's and 1980's. This class of models generates stochastic volatility of the short rate or yields based on a "level" effect in the underlying state variables. Denote an $ATSM$ with m state variables with square-root processes (which can be potentially correlated) and $n - m$ Gaussian state variables as an $A_m(n)$ (following the notation of Dai and Singleton (2000)). 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. Consequently, the $ATSM$ has two potential shortcomings in specifying stochastic volatility. First, out of n state variables, only m state varibles contribute to stochastic volatility

---

[2]Note that more broadly speaking, the term structure literature can be divided into two alternative approaches: the equilibrium approach and the arbitrage approach. The equilibrium approach is, as its name implies, a scheme to endogenously determine the term structure of the interest rate by specifying the 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 the market clearing condition. In contrast, the arbitrage approach relies upon extracting the information regarding pricing from the observed term structure 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 term structure of interest rate dynamics.

[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, Das, Foresi and Sundaram (1996), Chen (1996), Anderson and Lund (1997), and Jegadeesh and Pennachi (1996).

of the interest rate. Second, the variance of the interest rate is proportional to an affine function of the state variables themselves. The first problem cannot be solved trivially by equating n and m. As noted in Dai and Singelton (2000) and stressed in Ahn, Dittmar and Gallant (2000), while $n = m$ maximizes the *ATSM*'s flexibility in specifying heteroskedastic volatility, it limits its flexibility in specifying conditional/unconditional correlations among the state variables.

Compared to the *ATSM*s, non-affine term structure models have been relatively slow to develop. The representative non-affine term structure models are *QTSM*s. This class of models was first implied by Longstaff (1989) and pioneered by Beaglehole and Tenney (1991, 1992) and Constantinides (1992). However, only recently have these diverse versions been consolidated and systematically compared in a general framework. ADG formally define the *QTSM*s and also build a maximally flexible *QTSM* which can be identified empirically. The authors clarify the exact restrictions which reduce the general *QTSM* to existing sub-class models. The *QTSM* has the potential to be a more flexible model in specifying heteroskedacity in volatility since all included state variables can contribute to generating stochastic volatility unlike the *ATSM*. ADG also demonstrate that the *QTSM* is free of the trade-off between heteroskedastic volatilies of the interest rates and negative correlation among 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. Even though there is no strict one-to-one correspondence between the CIR factor in *ATSM*s and the Gaussian factor in *QTSM*s, we can easily demonstrate that the CIR factor is equivalent to the square of the Gaussian factor in its contribution to the volatility.[4] Thus in terms of generating volatility, $A_n(n)$ is equivalent to $Q(n)$.[5]

The last class which could be of importance 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. This model is unique in the sense that 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 $3/2$. This model 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.

In this paper, we investigate whether a mixture of these three different classes of models can enhance the performance of pure 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

---

[4]Using this concept, ADG formally proves that a particular version of the CIR model coincides the constrained version of *QTSM* with certain relevant restrictions.

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

heterogeneous parametric models, particularly in specifying the volatility of interest rates. If the hybrid model can improve the performance of parametric models in volatility specification, there might be an offsetting cost, which is of central importance in this paper. None of these questions have been pursued in the extant literature. We denote a hybrid model with m CIR 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 ADG (2000), we simultaneously use time series data on short- and long-term Treasury bond yields to explore the empirical properties of $A_m(n)Q(h)I(k)$. The hybrid models are of essentially two classes, in which $h = 2$ or $n = 2$. In the first case, we investigate whether augmenting a (correlated) quadratic two-state variable process with an (uncorrelated) affine (Gaussian or square-root process) or inverted square root process improves on our specification of volatility. Symmetrically, we also investigate whether an affine two-state variable process augmented by an uncorrelated affine or inverted square root process can improve the performance of the affine model.

We find several interesting results. First, we find that the specification of a maximal three-factor $QTSM$ ($A_0(0)Q(3)I(0)$) is very difficult to improve upon. However, by augmenting a two-factor $QTSM$ with an inverted-square root factor ($A_0(0)Q(2)I(1)$), we are able to improve upon the fit of the $QTSM$. Our diagnostics suggest that the model improves primarily on the ability of the $QTSM$ to fit the conditional mean of interest rates. However, like the maximal $QTSM$, the hybrid model is unable to fully capture the volatility dynamics. We find that augmenting various two-factor $ATSM$s by one-factor $QTSM$s result in dramatic improvement in the fit of the models relative to the maximal three-factor $ATSM$. However, none of these models are able to approach the performance of the maximal $QTSM$ or the $A_0(0)Q(2)I(1)$. In general, the cost-benefit tradeoff that emerges among the hybrid models appears to be that the models sacrifice the ability to fit the conditional mean of term structure dynamics in favor of conditional volatility, or vice versa.

The paper is organized as follows. In Section 2, we investigate alternative parametric models, $ATSM$s, $QTSM$s, and $ISRM$s focusing on their implications for heteroskedastic volatility. Section 3 provides a discussion of the 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 and their implication for heteroskedastic volatility are provided in Section 4, and Section 5 concludes.

2 Models

We define an economy represented by the augmented filtered probability space $(\Omega, F, \mathcal{F}, P)$, where $\mathcal{F} = \{\mathcal{F}_t\}_{0 \leq t \leq \mathcal{T}}$. First, we assume the existence of a stochastic discount factor, $M(t)$, which defines the canonical valuation equation:

$$x(t) = E_t^P \left[\frac{M(T)}{M(t)} x(T) \right] \tag{1}$$

where $x(t)$ is the price of an asset, with $x(t, \omega) : [0, \infty) \times \Omega \to R^+$. Absent arbitrage opportunities, there exists a unique equivalent probability measure, Q. Define a money market account $B(t)$ as

$$B(t) = e^{\int_0^t r(s)ds}.$$

As shown by Harrison and Kreps (1979) and Harrison and Pliska (1981), the scaled asset price $x(t)/B(t)$, is a martingale under the probability measure Q, i.e.

$$x(t) = E_t^Q \left[\frac{B(t)}{B(T)} x(T) \right] = E_t^Q \left[e^{-\int_t^T r(s)ds} x(T) \right] \tag{2}$$

In particular, for a discount bond which pays \$1 at maturity, the bond price is given as:

$$P(t, T) = E_t^Q \left[e^{-\int_t^T r(s)ds} \right] \tag{3}$$

In this paper, we investigate the performance of various term structure models under the assumption that the short term interest rate is a function of N unobserved state variables $Y(t) = (Y_1(t), Y_2(t), \cdots, Y_N(t))$ adapted to the probability space $(\Omega, F, \mathcal{F}, P)$. In particular we confine our attention to term structure models that are defined by the specified economy and the following assumptions:

Assumption 1: *The nominal instantaneous interest rate is a function of the state variables:*

$$r(t) = \delta_0 + \delta_A' Y_A(t) + \delta_Q' Y_Q(t) + \delta_I' Y_I(t) + Y_Q(t)' \Psi Y_Q(t) \tag{4}$$

where δ_0 is a constant; δ_A, δ_Q, and δ_I are n, h, and k dimensional vectors with $N = n + h + k$; Ψ is an $h \times h$ matrix of constants. A, I, and Q represent affine, inverted square root, and quadratic factors respectively.

Thus defined, the instantaneous interest follows a hybrid model of affine, quadratic and inverted square root processes, of which the stochastic differential equations are defined in the following assumptions:

Assumption 2: *The SDEs of the state variables $Y_A(t)$ are defined by a multivariate affine process:*

$$dY_A(t) = \mathcal{K}_A(\Theta_A - Y_A(t))dt + \Sigma_A \sqrt{S_A(t)} dz_A(t), \tag{5}$$

where $z_A(t)$ is an N_A-dimensional vector of independent Wiener processes, Θ_A is an n-dimensional vector, and $\mathcal{K}_A$ and Σ_A are $n \times n$ matrices. $S_A(t)$ is a diagonal matrix with the ith diagonal element given by $\alpha_{A,i} + \beta'_{A,i} Y_A(t)$, where $\beta_{A,i}$ is an n vector of constants.

Assumption 2 is taken directly from Dai and Singleton (2000).

Assumption 3: *The SDEs of the state variables $Y_Q(t)$ are defined by multivariate Gaussian processes:*

$$dY_Q(t) = \mathcal{K}_Q(\Theta_Q - Y_Q(t))dt + \Sigma_Q dz_Q(t), \tag{6}$$

where $z_Q(t)$ is an h-dimensional vector of independent Wiener processes, Θ_Q is an h-dimensional vector, and $\mathcal{K}_Q$ and Σ_Q are $h \times h$ matrix. Furthermore $\mathcal{K}_Q$ has non-negative eigenvalues.

Assumption 3 is taken from from Ahn, Dittmar and Gallant (2000), who specify the factor dynamics as a Gaussian processes.

Assumption 4: *The SDEs of the state variables $Y_I(t)$ are defined by independent multivariate inverted square root processes:*

$$dY_I(t) = \mathcal{K}_I(\Theta_I - Y_I(t)) \circ Y_I(t)dt + \Sigma_I \sqrt{S_I(t)}dz_I(t), \tag{7}$$

where $z_I(t)$ is a k-dimensional vector of independent Wiener processes, Θ_I is an k-dimensional vector. $\mathcal{K}$ and Σ_I are $k \times k$ diagonal matrices. $S_I(t)$ is also a $k \times k$ diagonal matrix with the ith element to be $Y_{I,i}(t)^{\frac{3}{2}}$. $\circ$ is a Hadamard product, an element by element multiplication.
This assumption is an extension of Ahn and Gao (2000) into a multi-variate setting with the assumption of mutually independent state variables.

In the following subsections, we specify each of the models that we estimate and discuss their implications for heteroskedastic volatility. As in Dai and Singleton (2000) and Ahn, Dittmar and Gallant (2000), 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 the factors is necessary to describe 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 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. The only exception is *ISRM*, for which an analytical bond price is not available for even two correlated factors.

7

2.1 Three Factor Quadratic Model: $A_0(0)Q(3)I(0)$

This is the three factor model studied by Ahn, Dittmar and Gallant (2000). The advantage of this model is its total flexibility in specifying the correlation structure among the factors. Since the model results in yield volatility that is a function of all of the factors, which are stochastic, the model has the potential to explain heteroskedasticity in volatility. However, Ahn, Dittmar and Gallant (2000) show that, although the model performs better than extant ones, it still has some defficiencies, particularly in capturing the observed volatility of yields. We employ this model as the benchmark case.

The model is governed by three quadratic factors following the risk-neutral process

$$dY(t) = \left[\begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} + \begin{pmatrix} A_{11} & 0 & 0 \\ A_{21} & A_{22} & 0 \\ A_{31} & A_{32} & A_{33} \end{pmatrix} Y(t) \right] dt + \begin{pmatrix} \sigma_{11} & 0 & 0 \\ 0 & \sigma_{22} & 0 \\ 0 & 0 & \sigma_{33} \end{pmatrix} d\tilde{Z}(t) \qquad (8)$$

the physical process

$$dY(t) = (\Phi_0 + \Phi_1 Y(t))dt + \Sigma dZ(t) \qquad (9)$$

with bond prices given by

$$P(t,\tau) = \exp\left[A(\tau) + B(\tau)'Y(t) + Y(t)'C(\tau)Y(t) \right] \qquad (10)$$

where the coefficients satisfy the ODEs

$$\begin{aligned}
\frac{dC(\tau)}{d\tau} &= 2C(\tau)\Sigma\Sigma'C(\tau) + \left(C(\tau)A + A'C(\tau)\right) - \mathbf{R} \\
\frac{dB(\tau)}{d\tau} &= 2C(\tau)\Sigma\Sigma'B(\tau) + A'B(\tau) + 2C(\tau)b - \beta \\
\frac{dA(\tau)}{d\tau} &= tr\left[\Sigma\Sigma'C(\tau)\right] + \frac{1}{2}B(\tau)'\Sigma\Sigma'B(\tau) + B(\tau)'b - \alpha
\end{aligned} \qquad (11)$$

As noted above, this specification allows for heteroskedastic volatility in bond yields; the diffusion term in the SDE for bond yields is a first-order function of the factors. Further, the framework allows for correlation among all of the factors. Consequently, for a model to better capture the dynamics of bond yields, the cost of omitting the correlations between the quadratic factors and the alternative factor must be outweighed by some additional flexibility in defining heteroskedastic volatility.

2.2 Two Quadratic and One Affine (Gaussion) Factor Model: $A_0(1)Q(2)I(0)$

The model to be considered here has two correlated quadratic factors and one independent Gaussian factor. This model is expected to perform worse than quadratic model since the independence

assumption restricts our ability to model correlations and the constant conditional volatility in the Gaussian process impairs our ability to fit heteroskedastic volatility.

The risk-neutral processes for the state variables are given as follows:

$$dY(t) = \left[\begin{pmatrix} b_1 \\ b_2 \\ \tilde{\kappa}\tilde{\theta} \end{pmatrix} + \begin{pmatrix} A_{11} & 0 & 0 \\ A_{21} & A_{22} & 0 \\ 0 & 0 & -\tilde{\kappa} \end{pmatrix} Y(t) \right] dt + \begin{pmatrix} \sigma_{11} & 0 & 0 \\ 0 & \sigma_{22} & 0 \\ 0 & 0 & \sigma \end{pmatrix} d\tilde{Z}(t) \tag{12}$$

and the physical process

$$dY(t) = \left[\begin{pmatrix} \Phi_{0,1} \\ \Phi_{0,2} \\ \kappa\theta \end{pmatrix} + \begin{pmatrix} \Phi_{1,11} & 0 & 0 \\ \Phi_{1,21} & \Phi_{1,22} & 0 \\ 0 & 0 & -\kappa \end{pmatrix} Y(t) \right] dt + \begin{pmatrix} \sigma_{11} & 0 & 0 \\ 0 & \sigma_{22} & 0 \\ 0 & 0 & \sigma \end{pmatrix} dZ(t) \tag{13}$$

with prices given by

$$P(t,\tau) = P_q(t,\tau)P_a(t,\tau) \tag{14}$$

where P_q represents the price for the two quadratic factors given by the solution to (10) and (11), and P_a is given by the Vasicek (1977) bond price formula.

2.3   Two Quadratic and One Affine (Square Root) Factor Model: $A_1(1)Q(2)I(0)$

With two quadratic and one square-root factor, this model is also expected to underperform the quadratic model. Although the square-root specification introduces level dependence of volatility on the factor, the effect is weaker than the level dependency afforded by quadratic model (the order of the effect is 0.5 in the square root case vs. 1.0 in the quadratic case). The loss of correlation between the square root factor and the quadratic factors will also hurt the model's performance.

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

$$dY(t) = \left[\begin{pmatrix} b_1 \\ b_2 \\ \tilde{\kappa}\tilde{\theta} \end{pmatrix} + \begin{pmatrix} A_{11} & 0 & 0 \\ A_{21} & A_{22} & 0 \\ 0 & 0 & -\tilde{\kappa} \end{pmatrix} Y(t) \right] dt + \begin{pmatrix} \sigma_{11} & 0 & 0 \\ 0 & \sigma_{22} & 0 \\ 0 & 0 & \sigma\sqrt{Y_3(t)} \end{pmatrix} d\tilde{Z}(t) \tag{15}$$

and the physical process

$$dY(t) = \left[\begin{pmatrix} \Phi_{0,1} \\ \Phi_{0,2} \\ \kappa\theta \end{pmatrix} + \begin{pmatrix} \Phi_{1,11} & 0 & 0 \\ \Phi_{1,21} & \Phi_{1,22} & 0 \\ 0 & 0 & -\kappa \end{pmatrix} Y(t) \right] dt + \begin{pmatrix} \sigma_{11} & 0 & 0 \\ 0 & \sigma_{22} & 0 \\ 0 & 0 & \sigma\sqrt{Y_3(t)} \end{pmatrix} dZ(t) \tag{16}$$

with prices given by

$$P(t,\tau) = P_q(t,\tau)P_a(t,\tau) \tag{17}$$

where P_q represents the prices for the two quadratic factors given by the solution to (10) and (11), and P_a is given by the CIR (1985) bond price formula.

2.4 Two Quadratic and One Inverted Square-Root Factor Model: $A_0(0)Q(2)I(1)$

Of the models considered in this section, this is the only model that offers potential trade-offs between correlation and level dependence. By replacing a correlated quadratic factor with an independent inverted-square-root factor, we loose flexibility in correlation structure. However, as strongly evidenced by Chan, Karolyi, Longstaff, and Sanders (1992), Aït-Sahalia (1996a, 1996b), 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 vs. 1.0), 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

$$
d\mathbf{Y}(t) = \left[\begin{pmatrix} b_1 \\ b_2 \\ \tilde{\kappa}\tilde{\theta} \end{pmatrix} + \begin{pmatrix} A_{11} & 0 & 0 \\ A_{21} & A_{22} & 0 \\ 0 & 0 & -\tilde{\kappa} \end{pmatrix} \mathbf{Y}(t) \right] dt + \begin{pmatrix} \sigma_{11} & 0 & 0 \\ 0 & \sigma_{22} & 0 \\ 0 & 0 & \sigma\sqrt{Y_3(t)^3} \end{pmatrix} d\tilde{Z}(t) \tag{18}
$$

and the physical process

$$
d\mathbf{Y}(t) = \left[\begin{pmatrix} \Phi_{0,1} \\ \Phi_{0,2} \\ \kappa\theta \end{pmatrix} + \begin{pmatrix} \Phi_{1,11} & 0 & 0 \\ \Phi_{1,21} & \Phi_{1,22} & 0 \\ 0 & 0 & -\kappa \end{pmatrix} \mathbf{Y}(t) \right] dt + \begin{pmatrix} \sigma_{11} & 0 & 0 \\ 0 & \sigma_{22} & 0 \\ 0 & 0 & \sigma\sqrt{Y_3(t)^3} \end{pmatrix} d\mathbf{Z}(t) \tag{19}
$$

with prices given by

$$
P(t,\tau) = P_q(t,\tau)P_I(t,\tau) \tag{20}
$$

where P_q represents the prices for the two quadratic factors given by the solution to (10) and (11), and P_I is given by the Ahn and Gao (1999) bond price formula

$$
P_I = \frac{\Gamma(\beta - \gamma)}{\Gamma(\beta)} M(\gamma, \beta, -x(Y_3(t), \tau)) x(Y_3(t), \tau)^\gamma \tag{21}
$$

where

$$
\begin{aligned}
x(Y_3(t), \tau) &= \frac{2(\kappa\theta - \lambda_1)}{\sigma^2 \left(e^{(\kappa\theta - \lambda_1)\tau} - 1 \right) Y_3(t)} \\
\beta &= \frac{2}{\sigma^2} \left[\kappa + \lambda_2 + (1 + \gamma)\sigma^2 \right] \\
\phi &= \kappa + \lambda_2 + \frac{1}{2}\sigma^2 \\
\gamma &= \frac{1}{\sigma^2} \left[\sqrt{\phi^2 + 2\sigma^2} - \phi \right]
\end{aligned}
$$

and $M(\cdot, \cdot, \cdot)$ is the confluent hypergeometric function.

2.5 Three Affine (Two Square-Root/One Gaussian) Factor Model: $A_2(3)Q(0)I(0)$

This is the maximal Dai and Singleton (2000) model. As demonstrated in Ahn, Dittmar, and Gallant (2000), this model fits the term structure dynamics more poorly than the quadratic model. Any dependence of volatility on the level of the state variables is weak compared to the quadratic or inverted square root case. This weaker dependence is due to the lower order level dependence of volatility in the square-root factors and constant volatility in the Gaussian factor.

Dai and Singleton (2000) define this model with two square-root and one Gaussian factor that follow the risk-neutral process

$$
d \begin{pmatrix} Y_1(t) \\ Y_2(t) \\ Y_3(t) \end{pmatrix} = \begin{pmatrix} \tilde{\kappa}_{11} & 0 & 0 \\ \tilde{\kappa}_{21} & \tilde{\kappa}_{22} & 0 \\ 0 & 0 & \tilde{\kappa}_{33} \end{pmatrix} \begin{pmatrix} \tilde{\theta}_1 - Y_1 \\ \tilde{\theta}_2 - Y_2 \\ -Y_3 \end{pmatrix} dt
$$
$$
+ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \sigma_{31} & 0 & 1 \end{pmatrix} \begin{pmatrix} \beta_{11}\sqrt{Y_1(t)} & 0 & 0 \\ 0 & \beta_{22}\sqrt{Y_2(t)} & 0 \\ 0 & 0 & \sqrt{Y_1(t)} \end{pmatrix} d\tilde{Z}(t) \tag{22}
$$

and the physical process

$$
dY(t) = \mathcal{K}\left(\Theta - Y(t)\right) dt + \Sigma S dZ(t) \tag{23}
$$

The bond price is given by

$$
P(t,\tau) = \exp\left[A(\tau) - B(\tau)'Y(t)\right] \tag{24}
$$

where the coefficients satisfy the ODEs

$$
\frac{dA(\tau)}{d\tau} = -\tilde{\Theta}'\tilde{\mathcal{K}}'B(\tau) + \frac{1}{2}\sum_{i=1}^{N}\left[\Sigma'B(\tau)\right]_i^2 \alpha_i - \delta_0
$$

$$
\frac{dB(\tau)}{d\tau} = -\tilde{\mathcal{K}}'B(\tau) - \frac{1}{2}\sum_{i=1}^{N}\left[\Sigma'B(\tau)\right]_i^2 \beta_i - \delta_y
$$
(25)

The coefficients δ are governed by the short rate equation;

$$
r(t) = \delta_0 + \delta_1 Y_1(t) + \delta_2 Y_2(t) + \delta_3 Y_3(t)
$$

where $\delta_1 = 0$, $\delta_2 = \delta_3 = 1$, and $\delta_0 = 0$.

2.6 One Quadratic and Two Affine (Square-Root/Gaussian) Factor Model: $A_1(2)Q(1)I(0)$

The remaining models are defined with two affine factors and a single quadratic factor. Because these models limit flexibility in both correlation structures and volatility specification, the models

are not expected to outperform the quadratic model. However, we can gain some insight into the relative importance of specifying flexibility in correlations vs. level effects by comparing the performance of these models to the full-fledged affine model.

The model of this subsection is determined by one quadratic, one square-root, and one Gaussian factor following the risk-neutral process

$$
d \begin{pmatrix} Y_1(t) \\ Y_2(t) \\ Y_3(t) \end{pmatrix} = \begin{pmatrix} \tilde{\kappa}_{11} & 0 & 0 \\ 0 & \tilde{\kappa}_{22} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \tilde{\theta}_1 - Y_1 \\ -Y_2 \\ b_1 + A_{11}Y_3 \end{pmatrix} dt
$$

$$
+ \begin{pmatrix} 1 & 0 & 0 \\ \sigma_{21} & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \beta_{11}\sqrt{Y_1(t)} & 0 & 0 \\ 0 & \sqrt{Y_1(t)} & 0 \\ 0 & 0 & \sigma_{11} \end{pmatrix} d\tilde{Z}(t) \qquad (26)
$$

and a corresponding physical process. The bond price is given by the product of a two-factor affine bond price, as in (24) and a one-factor quadratic bond price, as in (10).

2.7 One Quadratic and Two Affine (Square-Root) Factor Model: $A_2(2)Q(1)I(0)$

The performance of this model is expected to be somewhat similar to the previous model. The correlation between the factors is governed through the mean reversion matrix. In this specification, one quadratic and two Square-Root factors follow the risk-neutral process

$$
d \begin{pmatrix} Y_1(t) \\ Y_2(t) \\ Y_3(t) \end{pmatrix} = \begin{pmatrix} \tilde{\kappa}_{11} & 0 & 0 \\ \tilde{\kappa}_{21} & \tilde{\kappa}_{22} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \tilde{\theta}_1 - Y_1 \\ \tilde{\theta}_2 - Y_2 \\ b_1 + A_{11}Y_3 \end{pmatrix} dt
$$

$$
+ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \beta_{11}\sqrt{Y_1(t)} & 0 & 0 \\ 0 & \beta_{22}\sqrt{Y_2(t)} & 0 \\ 0 & 0 & \sigma_{11} \end{pmatrix} d\tilde{Z}(t) \qquad (27)
$$

and a corresponding physical process. The bond price is given by the product of a two-factor affine bond price, as in (24) and a one-factor quadratic bond price, as in (10).

2.8 One Quadratic and Two Affine (Gaussian) Factor Model: $A_0(2)Q(1)I(0)$

This model sacrifices the most flexibility in terms of level effects on volatility, as only one factor contributes to heteroskedastic volatility. In this model, one quadratic and two Gaussian factors follow the risk-neutral process

$$
d \begin{pmatrix} Y_1(t) \\ Y_2(t) \\ Y_3(t) \end{pmatrix} = \begin{pmatrix} \tilde{\kappa}_{11} & 0 & 0 \\ \tilde{\kappa}_{21} & \tilde{\kappa}_{22} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \tilde{\theta}_1 - Y_1 \\ \tilde{\theta}_2 - Y_2 \\ b_1 + A_{11}Y_3 \end{pmatrix} dt
$$

$$+ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \alpha_1 & 0 & 0 \\ 0 & \alpha_2 & 0 \\ 0 & 0 & \sigma_{11} \end{pmatrix} d\tilde{Z}(t) \qquad (28)$$

and a corresponding physical process. The bond price is given by the product of a two-factor affine bond price, as in (24) and a one-factor quadratic bond price, as in (10).

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 utilize the data set of McCulloch and Kwon (1993). These data are sampled at a monthly frequency and cover the period December 1946 through February 1991. Although this sample omits the most recent data, we view its advantages as superior to its disadvantages. Since the data set accounts for coupon payments, we observe a zero-coupon term structure that is the object of interest for our analysis.

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 and 12-month T-Bill yields and the 10-year bond yield. These maturities are similar to those examined in comparable studies, e.g. Dai and Singleton (2000), Ahn, Dittmar and Gallant (2000). All of the yields are treated as such in estimation; we do not use a proxy for the short rate. 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 Figure 2, which shows that the sample period covers a wide range of interest rate regimes, from very low levels in the late 1940's and 1950's to the high rate regime of the early 1980's. 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 $QTSM$s is that even in a single factor case, the short rate is not a sufficient statistic for risk in the economy. In the presence of multiple state variables, this issue becomes more important. 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, 1996b)]. Recent econometric advances have allowed researchers to address both of these issues through the use of simulated method of moments techniques. We specifically employ the Efficient Method of Moments [Gallant and Tauchen (1996)] to estimate the parameters of the

models investigated in this paper. This methodology has been used to estimate parameters of the short rate diffusion in Andersen and Lund (1997), to investigate *ATSM*s in Dai and Singleton (2000), and to investigate *QTSM*s in ADG (2000).[6]

The EMM procedure can be thought of as a two-step process. The first step is fitting a consistent estimator of the conditional density of the observable data. Designate this approximation to the density as

$$\hat{f}_K(y_t|x_{t-1}, \theta) = \frac{f(x_{t-1}, y_t|\theta)}{f(x_{t-1}|\theta)} \tag{29}$$

where y_t denotes the current observation of the observed process, x_{t-1} denotes lags of the process, and θ denotes the parameter vector of the density approximation. We approximate this density using the semi-nonparametric (SNP) procedure of Gallant and Tauchen (1989). The procedure augments a Gaussian vector-autoregression (VAR) with the potential for ARCH innovations by a Hermite polynomial expansion to capture deviations from normality. Designating a demeaned transformation of y_t as $z_t = R^{-1}(y_t - \mu)$, the SNP approximation to the density is given by

$$
\begin{aligned}
h_K(z_t|x_{t-1}) &= \frac{f(z_t, x_{t-1})\phi(z_t)}{\int f(s, x_{t-1})\phi(s)ds} \\
f(z_t, x_{t-1}) &= \sum_{|\alpha|=0}^{K_z} \sum_{|\beta|=0}^{K_x} (a_{\alpha\beta} x_{t-1}^\beta) z_t^\alpha \\
\phi(z_t) &\sim N(0, I).
\end{aligned}
\tag{30}
$$

We fit a SNP model to the Treasury data using the procedure outlined in Gallant and Tauchen (1997). The authors suggest an upward fitting strategy in which the parameters of parts of the SNP model are tuned to minimize the Schwartz (1978) criterion (BIC) and then used as starting points for the fitting of the next part of the model. This method provides a fairly efficient way to fit the model. Our Schwartz-preferred fit is described by $\{L_\mu, L_r, K_z, I_z, K_x\} = \{1, 4, 4, 3, 0\}$. $L_\mu = 1$ implies that one lag of the data is sufficient to describe mean dynamics in the VAR and $L_r = 4$ suggests that a fourth-order ARCH process describes the innovations to the process. $K_z = 4$ suggests that a fourth-order Hermite polynomial captures deviations from normality, and $I_z = 3$ indicates that the interaction terms in the orders of the polynomial are suppressed. Finally $K_x = 0$ suggests that it is unnecessary to incorporate lags of the process in modeling the coefficients of the Hermite polynomial. This specification is quite similar to that of other SNP specifications in term structure studies. For example, Dai and Singleton (2000) find a specification of $\{L_\mu, L_r, K_z, K_x\} = \{1, 2, 4, 0\}$ describes a term structure of six-month LIBOR, two-year swap, and ten-year swap yields over the period 1987-1996. Our specification differs only in the ARCH term, which likely reflects our incorporation of an earlier period in our data sample. ADG find that the same specification of the

---

[6]A detailed discussion of the method in these contexts can be found in Andersen and Lund (1997).

score generator describes the term structure of 3-month and 12-month Treasury Bills and 10-Year Treasury Bonds.

The second step in the EMM process involves estimating a parameter vector for the term structure model. The procedure takes a set of initial starting values for the model and simulates a long set of data. In our case, we set the simulation length to $T = 50,000$. The SNP model is fit to the simulated data and the scores of the fitted model with respect to the SNP parameters are estimated. Designate the parameters of the structural model as ρ and the parameters of the SNP model as $\tilde{\theta}$. The scores of the fitted SNP model are used as moment conditions, $m'(\rho, \tilde{\theta})$ and the quadratic form

$$m'(\rho, \tilde{\theta})\tilde{\mathcal{I}}^{-1}m(\rho, \tilde{\theta}) \tag{31}$$

is estimated, where $\tilde{\mathcal{I}}^{-1}$ denotes the quasi-information matrix from quasi-maximum likelihood estimation of θ. The procedure is repeated until the quadratic form is minimized. Then, a test of the specification of the SDE is formed through the test statistic

$$Tm'(\rho, \tilde{\theta})\tilde{\mathcal{I}}^{-1}(\rho, \tilde{\theta}) \sim \chi^2_{K-P} \tag{32}$$

where K denotes the dimension of θ and P denotes the dimension of ρ. The method uses all of the relevant moments of the conditional distribution and is therefore asymptotically as efficient as maximum likelihood, as shown in Gallant and Long (1997).

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 (1986)] and qualitative analysis using the reprojection methodology of Gallant and Long (1997).

4.1   EMM Specification Tests

Results of specification tests for four models are presented in Table 1. The models grouped in this table include at least two quadratic factors and are, from left to right, $A_0(0)Q(3)I(0)$, $A_0(1)Q(2)I(0)$, $A_1(1)Q(2)I(0)$, and $A_0(0)Q(2)I(1)$. The three-factor $QTSM$ is presented as a benchmark case. The results indicate that the $A_0(0)Q(3)I(0)$ performs fairly well in fitting the term structure dynamics; its ability to fit the dynamics is indicated by a z-statistic for the model specification of 6.976. While this result indicates statistical rejection of the model, it is similar to the results of ADG (2000), who find that $A_0(0)Q(3)I(0)$ outperforms $A_2(3)Q(0)I(0)$ in fitting a similar set of yields.

The second column presents results for a specification with two quadratic factors augmented by a Gaussian factor, $A_0(1)Q(2)I(0)$. The results indicate that replacing a third correlated quadratic factor with a Gaussian factor comes at a large cost in terms of fit; the z-statistic for the $A_0(1)Q(2)I(0)$ rises to 14.930, roughly double that of the $A_0(0)Q(3)I(0)$. The table suggests that the quadratic parameters remain largely unchanged; only in parameters that are poorly identified in the $A_0(0)Q(3)I(0)$ case are changes substantial. For example, the risk-neutral mean reversion coefficient $A(1,1)$ is estimated as -2.05 in the purebred model, but as -0.18 in the hybrid model. However, in neither case are the coefficients precisely estimated. The results suggest that the substitution of a Gaussian factor for a quadratic factor result in little impact on the form of the model and a reduction in the fit.

The results in the third column represent the specification of two quadratic factors augmented by a square-root factor, $A_1(1)Q(2)I(0)$. This specification actually performs worse than the $A_0(1)Q(2)I(0)$ specification, with a z-statistic of 18.977. This result is somewhat surprising, since the introduction of a square root factor improves the flexibility of the model in capturing heteroskedastic volatility. One of the main problems with this specification appears to be identification of the parameters; the majority of the parameters are imprecisely estimated; however, the source of this imprecision is unclear. Regardless, the results indicate that accommodating heteroskedastic volatility does not improve the performance of the model relative to $A_0(1)Q(2)I(0)$.

The final column presents results for a model with two quadratic factors and an inverse-square-root factor, $A_0(0)Q(2)I(1)$. The results indicate that this model performs somewhat better than the pure quadratic model, $A_0(0)Q(3)I(0)$, with the z-statistic falling to 6.393. The parameters of the model are estimated fairly precisely, with the exception of λ_1.[7] Thus, the results suggest that by augmenting the quadratic factors by a true nonlinear model, that the specification can better fit the dynamics of the term structure.

In the fifth column, we estimate another benchmark model, a three-factor quadratic model in which two factors are correlated and one factor is orthogonal. This estimation provides us greater insight as to the costs and benefits of specifying the correlation structure. The results indicate that restricting the correlation structure among the factors comes at a large cost; although the model has greater degrees of freedom, it is strongly rejected with a z-statistic of 23.507. We investigate the implications of this result in greater detail below.

To gain some better insights 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. These scores provide a diagnostic as to which features of the density that the models find difficult to fit. The terms $A(2) - A(13)$ represent the Hermite polynomial terms, and capture deviations from

---

[7] The results are not materially changed with λ_1 set to 0; this improves the z-statistic of the model by adding an additional degree of freedom.

conditional normality. The $\psi(1) - \psi(12)$ terms are the VAR terms in the model, representing the conditional mean of term structure dynamics. Finally, the $\tau(1) - \tau(78)$ terms represent the ARCH terms, capturing the conditional volatility of the term structure. The diagnostics for these scores are presented in Table 3.

The results suggest that a principal source of improvement in the $A_0(0)Q(2)I(1)$ model relative to the $A_0(0)Q(3)I(0)$ model is actually in the conditional mean of the process. None of the ψ terms for the former model exceed 1.96 in absolute value, suggesting that the scores with respect to these terms are fit fairly well by the hybrid model. In contrast, the $A_0(0)Q(3)I(0)$ model fails to match 6 of the 12 ψ scores. The models perform similarly in capturing the Hermite polynomial terms; each model misses two of the Hermite terms. The magnitude of the t-ratios for the $A_0(0)Q(3)I(0)$ model is considerably larger; however, both violate the scores at conventional significance levels. The results, however, indicate that both models should accommodate the shape of volatility reasonably well. Finally, along the dimension of the ARCH terms, the quadratic model fares somewhat better than the hybrid model, missing only two of the scores whereas the hybrid model misses four. Thus, the somewhat surprising result of the diagnostics is that the reason that the hybrid model performs better than the quadratic model is due to the conditional *mean* terms; both models appear to fit conditional volatility reasonably well, with the quadratic model perhaps having a slight advantage.

We also present diagnostics for the three-factor quadratic model with two correlated and one orthogonal factor. These results provide some greater insight into the comparison of the scores of the fully specified quadratic model and the $A_0(0)Q(2)I(1)$ model. The results suggest that restricting the correlation structure is very costly in terms of the volatility fit. The quadratic model with two correlated and one orthogonal factor performs somewhat similarly to the fully parameterized $A_0(0)Q(3)I(0)$ in terms of the Hermite polynomial terms and conditional mean terms. The model fails to fit one additional Hermite polynomial score and fails to fit the same conditional mean scores as the fully parameterized model. However, the model fares much more poorly in capturing the ARCH terms, violating 7 of these scores at conventional significance levels. These results indicate that the correlation structure is very important for matching the volatility dynamics observed in the data. Thus, the performance of the hybrid $A_0(0)Q(2)I(1)$ model is impressive given the restrictions imposed on the model by forcing the inverted square root factor to be orthogonal to the remaining factors.

In Table 2, we present results for models that have at least two affine factors; from left to right we present results for $A_0(2)Q(1)I(0)$, $A_1(2)Q(1)I(0)$, $A_2(2)Q(1)I(0)$, and $A_2(3)Q(0)I(0)$. The last model represents the preferred model of Dai and Singleton (2000). The results indicate that all four models perform quite poorly in describing term structure dynamics. The worst-performing model is the $A_2(2)Q(1)I(0)$ model, with a z-statistic of 63.511. However, it is interesting to note that two hybrid specifications improve upon the pure affine model; $A_0(2)Q(1)I(0)$ and $A_1(2)Q(1)I(0)$ outperform $A_2(3)Q(0)I(0)$. These results suggest that the flexibility in volatility specification

provided by the quadratic factor outweighs the flexibility in specifying conditional correlations. This is clearest in the case of the $A_0(2)Q(1)I(0)$ model, which performs the best of the four models in the table; the ability to model the heteroskedastic volatility through the Q term apparently provides a significant improvement to the fit of the model.

The diagnostics for the scores of these models are presented in Table 4. Three of the four models, $A_0(2)Q(1)I(0)$, $A_1(2)Q(1)I(0)$, and $A_2(3)Q(0)I(0)$ fit the conditional mean of the density quite well, with none of the t-ratios significant. In fact, the fit of the conditional means for these models appears generally superior to those of the models with multiple quadratic variables. However, the models fare much worse in terms of capturing conditional volatility and deviations from normality. Three of the models have significant scores for the Hermite terms; the fourth model, $A_2(3)Q(0)I(0)$ has significant scores for eight of the Hermite terms. The model that best captures the ARCH terms for the conditional volatility in the model is the $A_1(2)Q(1)I(0)$ model, which fails 7 of the ARCH scores. Thus, none of these models appears able to capture volatility and departure from normality particularly well.

In summary, the somewhat surprising result of our specification test was that the pure quadratic model continues to provide the best specification for volatility and departures from normality, with a two quadratic factor-model with one inverted square root factor following closely behind. Overall, this latter model appears to perform better than the quadratic model, but the source of the improvement appears to be in the specification of the conditional mean terms rather than in the volatility. This is somewhat surprising since the inverted square root factor provides some additional flexibility in modeling heteroskedastic volatility. In contrast, the affine-dominated hybrid models do not perform as well as the quadratic-dominated hybrid models. However, these models display some considerable improvement in fit relative to the pure affine case.

4.2   Reprojection

We briefly summarize the reprojection method in this section; a complete discussion is provided in Gallant and Tauchen (1998). We denote the conditional density implied by the $QTSM$ for observables as

$$\hat{p}\left(y_0 | y_{-L}, ..., y_{-1}\right) = \left(y_0 | y_{-L}, ..., y_{-1}, \hat{\rho}_n\right) \tag{33}$$

where $\hat{\rho}$ denotes the estimated model parameters. Although analytic expressions for (33) are not available, an unconditional expectation, $\mathcal{E}_{\hat{\rho}_n}$ can be computed by generating a simulation $\{\hat{y}_t\}_{t=-L}^N$ from the system with parameters set to $\hat{\rho}_n$ and approximating

$$\mathcal{E}_{\hat{\rho}_n}(g) = \frac{1}{N} \sum_{t=0}^N g(\hat{y}_{t-L}, \dots, \hat{y}_t) \tag{34}$$

With respect to unconditional expectation so computed, define

$$\hat{\theta}_K = \arg\max \theta \in \Re^{p_K} \; \mathcal{E}_{\hat{\rho}_n} \log f_K(y_0|y_{-L}, \ldots, y_{-1}, \theta) \tag{35}$$

where $f_K(y_0|y_{-L}, \ldots, y_{-1}, \theta)$ is the SNP conditional density (30). Let

$$\hat{f}_K(y_0|y_{-L}, \ldots, y_{-1}) = f_K(y_0|y_{-L}, \ldots, y_{-1}, \hat{\theta}_K). \tag{36}$$

Theorem 1 of Gallant and Long (1997) states that

$$\lim_{K \to \infty} \hat{f}_K(y_0|y_{-L}, \ldots, y_{-1}) = \hat{p}(y_0|y_{-L}, \ldots, y_{-1}).$$

Convergence is with respect to a weighted Sobolev norm that they describe. Of relevance here is that convergence in their norm implies that $\hat{f}_K$ as well as its partial derivatives in $(y_{-L}, \ldots, y_{-1}, y_0)$ converge uniformly over $\Re^\ell$, $\ell = M(L+1)$, to those of $\hat{p}$. The idea of reprojection is to study the dynamics of $\hat{p}$ by using $\hat{f}_K$ as an approximation; this result provides the justification.

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

$$\mathcal{E}(y_0|y_{-L}, \ldots, y_{-1}) = \int y_0 \, f_K(y_0|x_{-1}, \hat{\theta}_K) \, dy_0$$

and

$$\text{Var}(y_0|y_{-L}, \ldots, y_{-1})$$
$$= \int [y_0 - \mathcal{E}(y_0|x_{-1})] \, [y_0 - \mathcal{E}(y_0|x_{-1})]' \; f_K(y_0|x_{-1}, \hat{\theta}_K) \, dy_0$$

where $x_{-1} = (y_{-L}, \ldots, y_{-1})$. Thus the reprojection method provides some further insight into the performance of the models in capturing the conditional means, volatilitities, and deviations from normality implied by the data.

Plots of the first two conditional moments implied by six of the eight models are presented in Figures 2 through 7. The figures depict the conditional moments implied by the models (dashed lines) and the conditional moments implied by the score generator (solid lines). Figure 2 presents results for the fully specified three-factor quadratic model. The graphs of the conditional means suggests that the model's difficulty in fitting conditional means is largely concentrated in the short end of the term structure. The three-factor quadratic model performs quite well in capturing the conditional mean of the ten-year yield, performs somewhat worse for the twelve-month yield, and captures the conditional mean of the six-month yield fairly poorly. A similar pattern emerges for volatility; the model reproduces the time series pattern and the level of the 10-year volatility fairly well. However, the model cannot reproduce the level of conditional volatility found in the six-month and twelve-month yields.

The results for the hybrid model $A_0(0)Q(2)I(1)$ are presented in Figure 3. The graph suggests that, as indicated by the t-ratio diagnostics, that the model performs much better in capturing the

conditional mean of term structure dynamics. The model matches the dynamics of the twelve-month and 10-year yields quite well, with some slightly larger discrepancies for the twelve-month yield. The model has a slightly more difficult time matching the 6-month yield conditional mean; however, the deviations in the model conditional means and the means implied by the conditional distribution are not as large as for the pure quadratic model case. Conditional volatility, however, is not captured as well by this hybrid model. Again, the model tracks the time series and level of the ten-year yield volatility reasonably well. However, significant deviations are present for the shorter-maturity bonds, and the conditional volatility of the $A_0(0)Q(2)I(1)$ does not approach the level of volatility implied by the score generator.

The remaining figures suggest that models with multiple affine factors have severe difficulty in capturing many of the features of the data. The reprojections suggest that the models perform quite poorly in capturing both conditional means and conditional volatilities. Interestingly, the $A_2(3)Q(0)I(0)$ model displays the reverse pattern of the quadratic models; it fits the conditional mean of the long bond quite poorly and performs better in fitting the conditional mean of the shorter term bonds. Further, all of these models imply conditional volatilities that, relative to the conditional volatility implied by the score generator, are nearly flat.

In general, the results of the reprojection conform to the intuition provided by the test diagnostics. The main finding is that there appears to be some tradeoff in modeling term structure dynamics by incorporating a third (correlated) quadratic state variable compared to a single (uncorrelated) inverted square root state variable. The inverted square root variable appears to improve the ability of the quadratic model to capture conditional means of bond yields. However, this improvement comes at the cost of matching the conditional volatilities of these yields. As mentioned previously, this result seems somewhat surprising given the additional flexibility provided by the diffusion term of the inverted square root model.

5 Conclusion

This study confirms, outside of the affine factor specification, the claim made in Dai and Singleton (2000) of a trade-off between dependence of of the conditional variance of factors on factor levels and the admissible structure of the correlation matrix for the factors. In particular, we find that a hybrid quadratic/inverted-square-root model provides a better fit to term structure dynamics than the benchmark quadratic model, albeit marginally. This performance is achieved despite the more restrictive correlation structure implied by the hybrid model, by allowing but higher level dependence. The results of models with two affine factors and a quadratic factor appear to underscore the trade-off between level dependence and correlation structure; although none of these models fit the term structure dynamics well, the models improve upon the performance of the maximal affine model. This result may suggest that level dependence must be given similar weight

to correlation structure in specifying factor dynamics. We expect that a model that allows for a fully specified correlation structure would perform even better than the hybrid quadratic-inverted square root model, but such a study must rely on a numerical solution of the PDE for bond prices.

The study does leave us with some puzzles. On detailed examination of the hybrid model, we find that the improved performance appears to come from better fit in the conditional mean of the process rather than the volatility. Given the flexibility of the hybrid model in specifying level effects in volatility, this result appears somewhat counterintuitive. One possible explanation for this result is that correlations among factors are reflected in the yield volatility since we are only able to observe discrete realizations of the yields. Because the quadratic model offers a more flexible correlation structure, it may be better able to capture this effect despite its relative deficiency in level dependence. Regardless of the source, both of these effects appear to be important in capturing the moments of the term structure dynamics. Future progress in the specification of both volatility dynamics and correlation structure will allow these moments to better be captured by term structure models.

References

Ahn, D., Dittmar, R., and A. R. Gallant, "Quadratic Term Structure Models: Theory and Evidence," unpublished manuscript, University of North Carolina.

Ahn, D., and B. Gao, 1999, "A Parametric Nonlinear Model of Term Structure Dynamics," *Review of Financial Studies*, **12**, 721-762.

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. and J. Lund, 1997, "Estimating Continuous-time Stochastic Volatility Models of the Short-Term Interest Rate," *Journal of Econometrics*, **77**, 343-377.

Balduzzi, P., S. Das, S. Foresi, and R. Sundaram, 1996, "A Simple Approach to Three Factor Affine Models of the Term Structure," *Journal of Fixed Income*, **6**, 43-53.

Beaglehole, D., and M. Tenney, 1991, "General Solutions of Some Interest Rate-Contingent Claim Pricing Equations," *Journal of Fixed Income*, **1**, 69-83.

Beaglehole, D., and M. Tenney, 1992, "A Nonlinear Equilibrium Model of the Term Structure of Interest Rates: Corrections and Additions," *Journal of Financial Economics*, **32**, 345-454.

Chan, K. C., G. A. Karolyi, F. A. Longstaff, and A. B. Sanders, 1992, "An Empirical Comparison of Alternative Models of the Short-Term Interest Rate," *Journal of Finance*, **47**, 1209-1227.

Chen, R.R., and L. Scott, 1992, "Pricing Interest Options in a Two-Factor Cox-Ingersoll-Ross Model of the Term Structure," *Review of Financial Studies*, **5**, 613-636.

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.

Constantinides, G., 1992, "A Theory of the Nominal Structure of Interest Rates," *Review of Financial Studies*, **5**, 531-552.

Cox, J.C., J.E. Ingersoll, and S.A. Ross, 1985, "A Theory of the Term Structure of Interest Rates," *Econometrica*, **53**, 385-406.

Dai, Q., and K. Singleton, 200, "Specification Analysis of Affine Term Structure Models", *Journal of Finance*, **55**, 1943-1978.

Duffie, D. and R. Kan, 1996, "A Yield-Factor Model of Interest Rates," *Mathematical Finance*, **6**, 379-406.

Engle, R., and V. Ng, 1993, "Time-Varying Volatility and the Dynamic Behavior of the Term Structure," *Journal of Money, Credit, and Banking*, **25**, 336-349.

Gallant, A. R., and J. R., Long, 1997, "Estimating Stochastic Differential Equations Efficiently by Minimum Chi-Squared," *Biometrika*, **84**, 125-141.

Gallant, A. R., and G. E. Tauchen, 1996, "Which Moments to Match?" *Econometric Theory*, **12**, 657-681.

Gallant, A. R., and G. E. Tauchen, 1997, "SNP: A Program for Nonparametric Time Series Analysis," unpublished manuscript, Duke University.

Gallant, A. R., and G. E. Tauchen, 1998, "Reprojecting Partially Observed Systems with Application to Interest Rate Diffusions," *Journal of the American Statistical Association*, **93** 10-24.

Harrison, M., and D. Kreps, 1979, "Martingales and Arbitrage in Multiperiod Security Markets," *Journal of Economic Theory* **20** 381-408.

Harrison, M., and S. Pliska, 1981, "Martingales and Stochastic Integrals in the Theory of Continuous Trading," *Stochastic Processes and their Applications*, **11**, 215-260.

Jegadeesh, N. and G. Pennachi, 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.

Longstaff, F., 1989, "A Nonlinear General Equilibrium Model of the Term Structure of Interest Rates," *Journal of Finance*, **23** 1259-1282.

Longstaff, F., and E. Schwartz, 1992, "Interest Rate Volatility and the Term Structure: A Two-Factor General Equilibrium Model," *Journal of Finance* **47** 1259-1282.

McCulloch, J., and H. Kwon, 1993, "U.S. Term Structure Data, 1947-1991," unpublished manuscript, Ohio State University.

Pearson, N. D., and T. S. Sun, 1994, "Exploiting the Conditional Density in Estimating the Term Structure: An Application to the Cox, Ingersoll, and Ross Model," *Journal of Finance*, **49**, 1279-1304.

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.

Table 1: **Specification Tests of Term Structure Models**

Coefficient	$A_0(0)Q(3)I(0)$	$A_0(1)Q(2)I(0)$	$A_1(1)Q(2)I(0)$	$A_0(0)Q(2)I(1)$	$A_0(0)Q^{***}(3)I(0)$
α	0.0381 (0.0207)			0.0405 (0.0040)	
$R(2,1)$	-0.6411 (2.1362)	-0.5934 (0.4462)	-0.7162 (2.3059)	-0.2736 (0.0437)	
$R(3,1)$	0.4541 (1.5846)				
$R(3,2)$	-0.9571 (0.0889)				-0.7858 (0.2688)
$b(1)$	-0.0361 (1.0103)	-0.0322 (0.2676)	-0.0094 (0.0337)	0.1377 (0.0180)	0.1323 (0.3600)
$b(2)$	-0.0583 (0.1745)	-0.2772 (0.1574)	-0.0210 (0.0972)	-0.0400 (0.0045)	-0.0430 (0.0157)
$b(3)$	0.2214 (0.6868)				0.2202 (0.0772)
$A(1,1)$	-2.0490 (9.0499)	-0.1803 (0.8635)	-0.1245 (1.2738)	-1.6332 (0.1212)	-1.4798 (2.5891)
$A(2,1)$	0.0181 (1.8000)	0.1097 (0.0890)	-0.1480 (1.5970)	0.2882 (0.0299)	
$A(3,1)$	-2.6043 (6.5106)				
$A(2,2)$	-0.2470 (0.0415)	-0.2181 (0.0298)	-0.1103 (0.1407)	-0.1217 (0.0146)	-0.2373 (0.0540)
$A(3,2)$	1.2125 (0.5019)				1.3444 (0.3623)
$A(3,3)$	-0.2631 (0.0434)				-0.2779 (0.1007)
$\Sigma(1,1)$	1.5185*(0.3130)	2.3975*(0.9507)	0.4113*(2.8703)	14.0968* (1.8088)	1.9454* (7.6299)
$\Sigma(2,2)$	0.1559*(0.1382)	0.0629*(0.0135)	6.3587*(49.0821)	1.0240* (0.0596)	0.0629* (0.0352)
$\Sigma(3,3)$	0.5938*(0.2064)				0.7645* (0.1180)
$\Phi_0(1)$	-0.0411 (0.7745)	-0.1455 (0.0560)	1.0112 (2.9161)	0.0027 (0.0007)	-0.0430 (0.1148)
$\Phi_0(2)$	-0.0008 (0.0000)	0.0026 (0.0066)	0.0941 (0.0351)	-0.0002 (0.0000)	-0.0015 (0.0006)
$\Phi_0(3)$	0.1693 (0.7101)				0.1618 (0.0562)
$\Phi_1(1,1)$	-1.7637 (0.3621)	-1.1853 (0.4421)	-6.7895 (16.8038)	-0.8950 (0.1276)	-0.6192 (0.6202)
$\Phi_1(2,1)$	0.0039 (0.0352)	0.0001 (0.0627)	-0.6665 (0.7573)	0.0620 (0.0111)	
$\Phi_1(3,1)$	-2.7877 (0.0000)**				
$\Phi_1(2,2)$	-0.0038 (0.0006)	-0.0974 (0.1486)	-0.0051 (0.0029)	-0.0022 (0.0003)	-0.0108 (0.0024)
$\Phi_1(3,2)$	1.2091 (0.4214)				0.9735 (0.2994)
$\Phi_1(3,3)$	-0.3417 (0.0312)				-0.3166 (0.1323)
κ		0.0003 (0.0002)	0.5818 (0.2011)	2.8594 (2.0308)	
θ		-0.0035 (0.0140)	0.0439 (0.0037)	0.0023 (0.0010)	
σ		0.0359 (0.0139)	0.0035 (0.0042)	8.2801 (2.9192)	
λ_0				0.0000	
λ_1		0.0100 (0.0000)**	-15.6120 (0.0000)**	-0.1175**(0.0000)	
χ^2	57.677	130.570	159.185	68.291	186.859
df	17	25	25	24	24
z	6.976	14.930	18.977	6.393	23.507

$^* \times 10^{-4}$

** Coefficient not estimated precisely.

** First factor is orthogonal to second and third factors.

Table presents results of estimation of four term structure models. 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.

Table 2: **Specification Tests of Term Structure Models**

Coefficient	$A_0(2)Q(1)I(0)$	$A_1(2)Q(1)I(0)$	$A_2(2)Q(1)I(0)$	$A_2(3)Q(0)I(0)$
α	-0.0126	0.0286		
$b(1)$	-0.0650	0.1692	-0.2275	
$A(1,1)$	-1.8489	-1.5190	-2.1201	
$\Sigma(1,1)$	0.0137	8.3780**	1.5952**	
$\Phi_0(1)$	0.0654	0.1050	0.1476	
$\Phi_1(1,1)$	-1.7478	-1.0948	-1.1168	
$\kappa(1,1)$	0.0050	0.1309	0.3170	5.0232
$\kappa(2,1)$	-29.8649		0.0917	2.3304
$\kappa(2,2)$	10.6862	3.6785	4.1378	0.1282
$\kappa(3,3)$				7.7198
$\theta(1)$	-0.0073	0.0181	0.0135	0.0017
$\theta(2)$	-0.0537		0.0127	0.0777
$\alpha(1)$	0.0000			
$\alpha(2)$	0.0273			
$\beta(1,1)$		0.0002	0.0019	0.0013
$\beta(2,2)$		1.0000*	0.0016	0.0000
$\sigma(2,1)$		-70.9546		
$\sigma(3,1)$				0.8680
$\lambda(1)$	0.5364	0.0021	0.0607	0.1425
$\lambda(2)$	-0.0100	-0.0000	-0.0139	-1.6857
$\lambda(3)$				-4.2685
χ^2	246.524	296.663	503.270	409.936
df	27	29	28	30
z	29.873	35.146	63.511	49.050

*fixed
** $\times 10^{-4}$

Table presents results of estimation of four term structure models. 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.

Table 3: **EMM Diagnostics of Term Structure Models**

Coefficient	$A_0(0)Q(3)I(0)$	$A_0(1)Q(2)I(0)$	$A_1(1)Q(2)I(0)$	$A_0(0)Q(2)I(1)$	$A_0(0)Q^*(3)I(0)$
A(2)	-0.172	0.452	-0.468	-1.301	0.723
$A(3)$	-1.314	-0.358	-0.554	-1.276	-0.223
$A(4)$	0.025	0.240	0.540	0.137	-0.106
$A(5)$	0.215	1.935	2.648	1.132	0.302
$A(6)$	0.116	1.305	1.813	-0.092	1.535
$A(7)$	3.987	4.449	4.585	0.857	3.345
$A(8)$	-0.124	0.557	0.253	-0.635	0.368
$A(9)$	-1.206	0.053	0.279	-1.212	0.129
$A(10)$	-0.344	-0.434	-0.107	-0.724	-0.635
$A(11)$	1.157	3.713	4.307	2.431	1.773
$A(12)$	0.650	2.546	2.963	0.831	2.645
$A(13)$	4.506	4.683	4.547	2.205	3.949
$\psi(1)$	-1.739	0.751	-0.323	-0.442	-0.799
$\psi(2)$	1.209	-3.088	-1.563	-0.444	0.899
$\psi(3)$	-0.709	2.528	0.028	-1.071	0.605
$\psi(4)$	-2.177	-1.098	0.340	1.193	2.418
$\psi(5)$	-2.325	1.917	-0.586	-1.790	-2.735
$\psi(6)$	1.251	-2.060	0.043	1.248	1.251
$\psi(7)$	2.390	-1.175	0.376	1.403	2.678
$\psi(8)$	-2.457	1.897	-0.554	-1.949	-2.932
$\psi(9)$	1.367	-1.958	-0.002	1.339	1.251
$\psi(10)$	2.138	-0.774	0.476	1.453	2.084
$\psi(11)$	-2.189	1.380	-0.455	-1.899	-2.584
$\psi(12)$	1.275	-1.488	-0.225	1.294	1.060
$\tau(1)$	1.366	1.565	1.993	1.101	0.913
$\tau(2)$	-0.738	0.113	-2.238	-1.793	-2.735
$\tau(3)$	1.344	0.674	3.153	2.101	3.521
$\tau(4)$	-0.249	-2.608	-3.694	-2.118	-5.416
$\tau(5)$	0.376	2.706	4.309	2.632	6.159
$\tau(6)$	1.824	1.860	0.183	0.975	-5.577
$\tau(7)$	1.264	2.299	2.627	1.073	1.666
$\tau(15)$	1.129	0.788	1.113	1.305	1.556
$\tau(24)$	3.014	2.699	1.631	2.216	-0.909
$\tau(25)$	1.786	2.980	3.475	1.812	2.294
$\tau(33)$	0.079	0.095	1.121	0.823	1.732
$\tau(42)$	2.123	1.898	0.989	1.160	-0.789
$\tau(43)$	1.672	2.591	2.562	1.293	1.913
$\tau(51)$	1.034	0.719	1.601	1.663	1.930
$\tau(60)$	0.672	1.346	0.574	0.479	-1.992
$\tau(61)$	0.580	0.975	1.133	0.554	0.773
$\tau(69)$	1.205	0.214	1.348	1.107	1.423
$\tau(78)$	1.097	1.176	0.466	0.305	-1.147

* First factor is orthogonal to second and third factors.

Table presents diagnostics for EMM scores. The coefficients labeled $A(k)$ denote Hermite polynomial terms in the SNP score generator, $\Psi(k)$ denotes the AR terms in the VAR of the SNP score generator, and $\tau(k)$ denotes the ARCH terms in the VAR 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.

Table 4: **EMM Diagnostics of Term Structure Models**

Coefficient	$A_0(2)Q(1)I(0)^*$	$A_1(2)Q(1)I(0)$	$A_2(2)Q(1)I(0)$	$A_2(3)Q(0)I(0)^*$
A(2)	0.805	-0.659	1.540	-1.167
A(3)	-0.852	-0.828	0.418	-2.393
A(4)	0.103	-0.232	-1.213	-2.694
A(5)	1.981	3.320	2.125	2.089
A(6)	1.881	2.953	2.634	1.205
A(7)	3.133	4.672	2.284	-6.335
A(8)	-0.190	0.124	1.151	-0.161
A(9)	0.253	-0.022	0.628	-0.323
A(10)	-0.480	-0.712	-1.310	-4.298
A(11)	3.607	4.568	3.713	3.850
A(12)	2.688	2.569	2.948	2.394
A(13)	3.997	4.394	3.323	-6.585
$\psi(1)$	-0.331	-0.545	-1.727	-0.889
$\psi(2)$	-0.808	-1.567	2.456	0.493
$\psi(3)$	-0.320	-0.464	0.161	0.267
$\psi(4)$	-0.441	0.242	2.346	-0.485
$\psi(5)$	0.042	-0.471	-2.410	-0.023
$\psi(6)$	-0.186	0.568	0.695	0.334
$\psi(7)$	-0.082	0.248	2.418	-0.603
$\psi(8)$	-0.364	-0.406	-2.408	0.101
$\psi(9)$	-0.180	-0.523	0.588	0.341
$\psi(10)$	-0.272	0.220	1.936	-0.385
$\psi(11)$	-0.049	-0.064	-2.087	0.109
$\psi(12)$	-0.181	0.368	0.132	0.298
$\tau(1)$	0.915	2.366	0.439	-1.789
$\tau(2)$	-3.217	0.372	-4.001	-5.160
$\tau(3)$	4.272	1.794	5.769	5.723
$\tau(4)$	-5.102	-5.405	-5.698	-2.367
$\tau(5)$	4.923	5.507	4.610	2.469
$\tau(6)$	1.085	2.128	5.740	1.629
$\tau(7)$	1.520	1.922	0.914	-3.226
$\tau(15)$	2.257	0.589	2.721	2.958
$\tau(24)$	2.308	2.195	4.111	2.345
$\tau(25)$	2.450	2.648	1.783	-2.721
$\tau(33)$	2.636	0.382	3.681	4.012
$\tau(42)$	1.523	1.648	2.920	1.627
$\tau(43)$	1.914	2.111	1.397	-1.742
$\tau(51)$	2.804	0.523	3.445	3.759
$\tau(60)$	1.160	1.371	2.745	1.128
$\tau(61)$	0.766	0.906	0.518	-1.117
$\tau(69)$	1.925	0.456	2.376	2.323
$\tau(78)$	0.969	1.060	2.117	1.015

[*] Quasi-t ratios. Actual will probably be larger.

Table presents diagnostics for EMM scores. The coefficients labeled $A(k)$ denote Hermite polynomial terms in the SNP score generator, $\Psi(k)$ denotes the AR terms in the VAR of the SNP score generator, and $\tau(k)$ denotes the ARCH terms in the VAR 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.
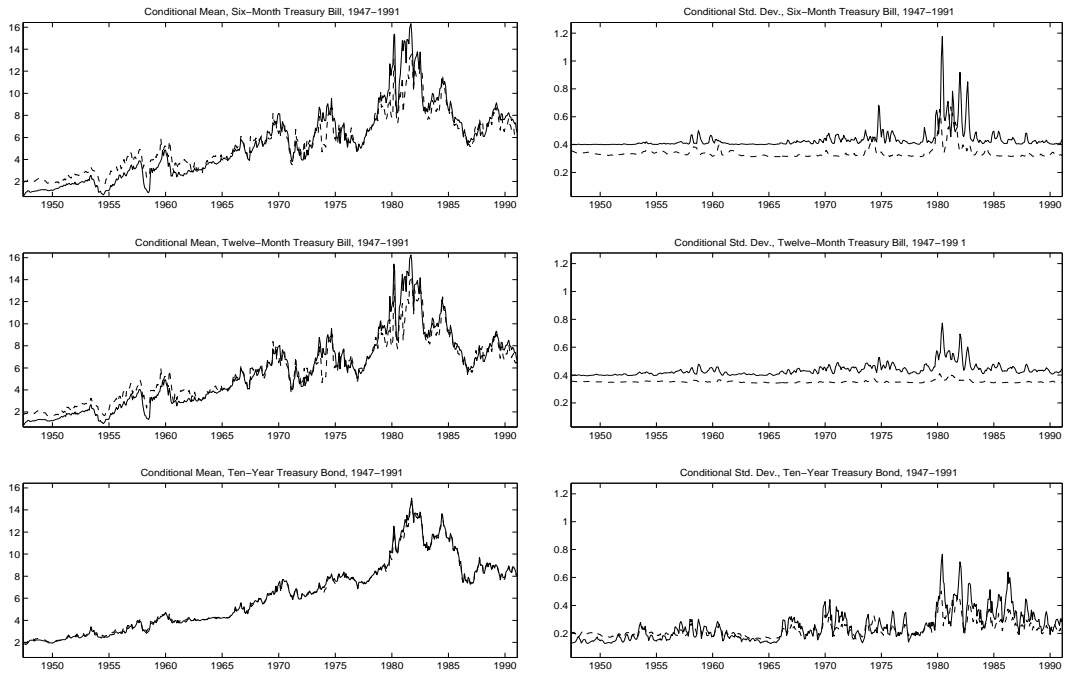
(a) Conditional Means   (b) Conditional Volatility
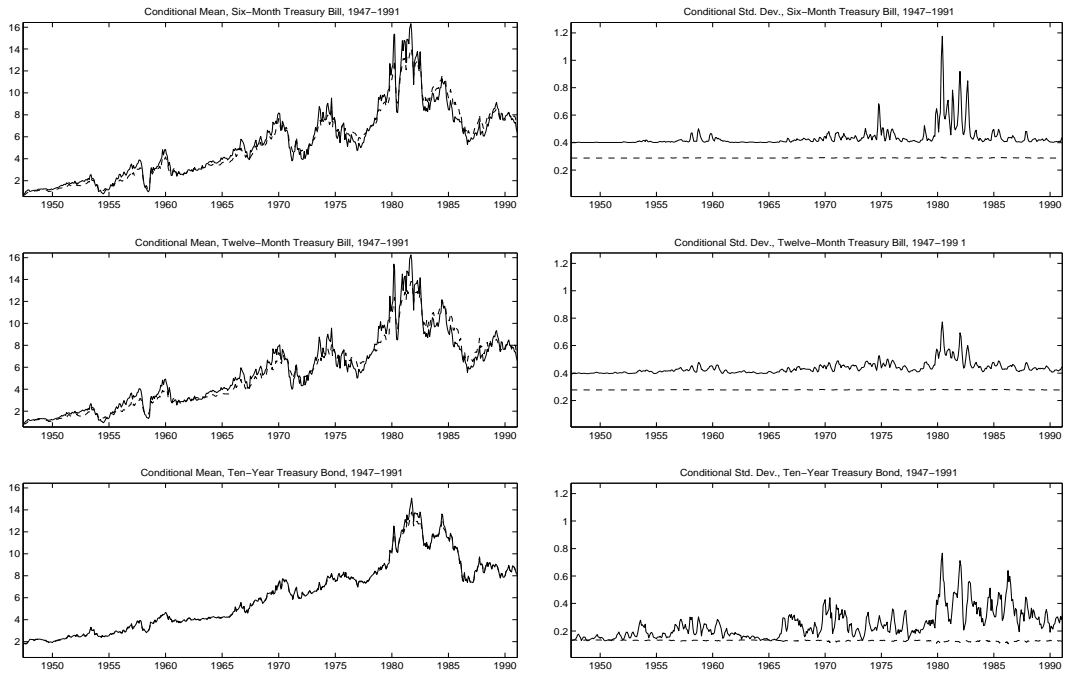
Figure 1: Reprojected Conditional Moments: $A_0(0)Q(3)I(0)$
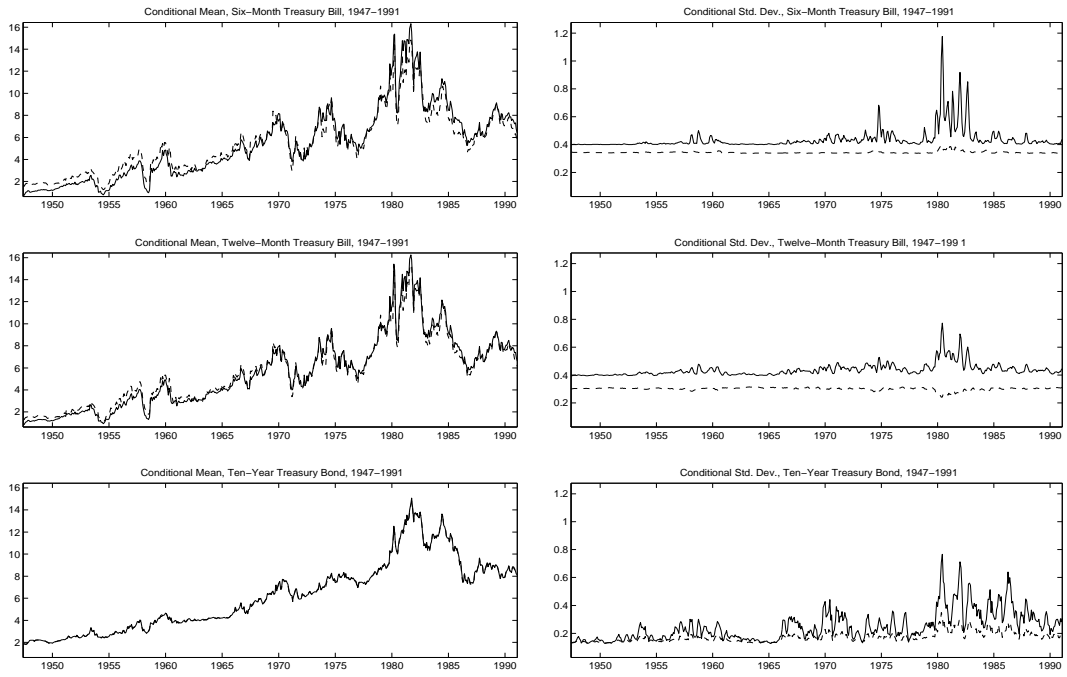
(a) Conditional Means   (b) Conditional Volatility
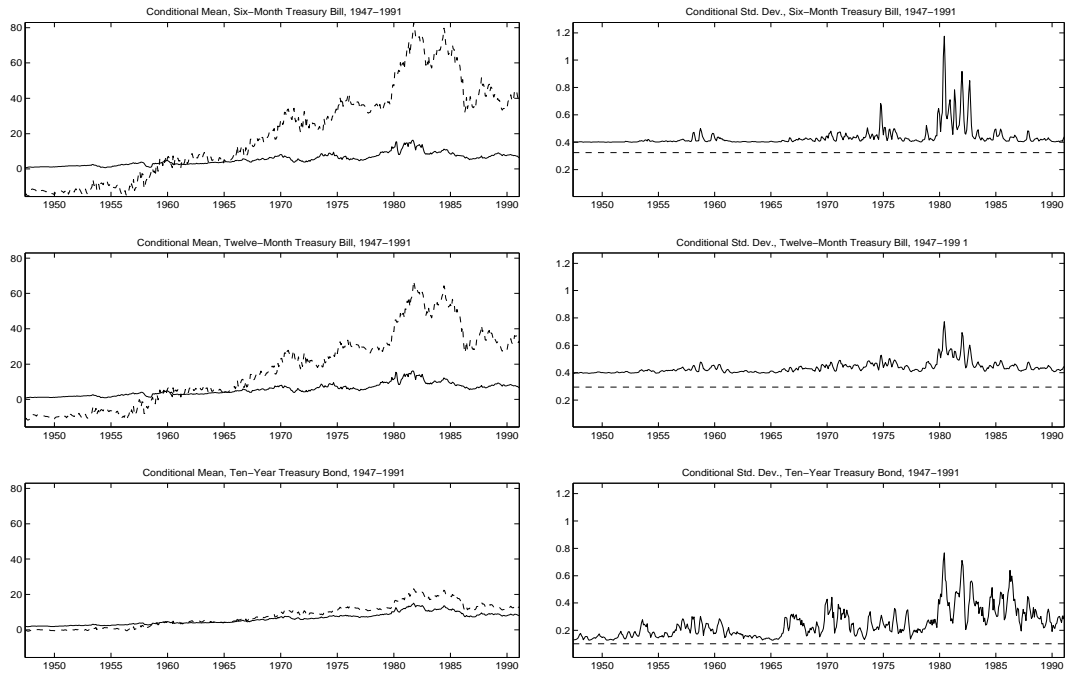
Figure 2: Reprojected Conditional Moments: $A_0(0)Q(2)I(1)$

(a) Conditional Means   (b) Conditional Volatility

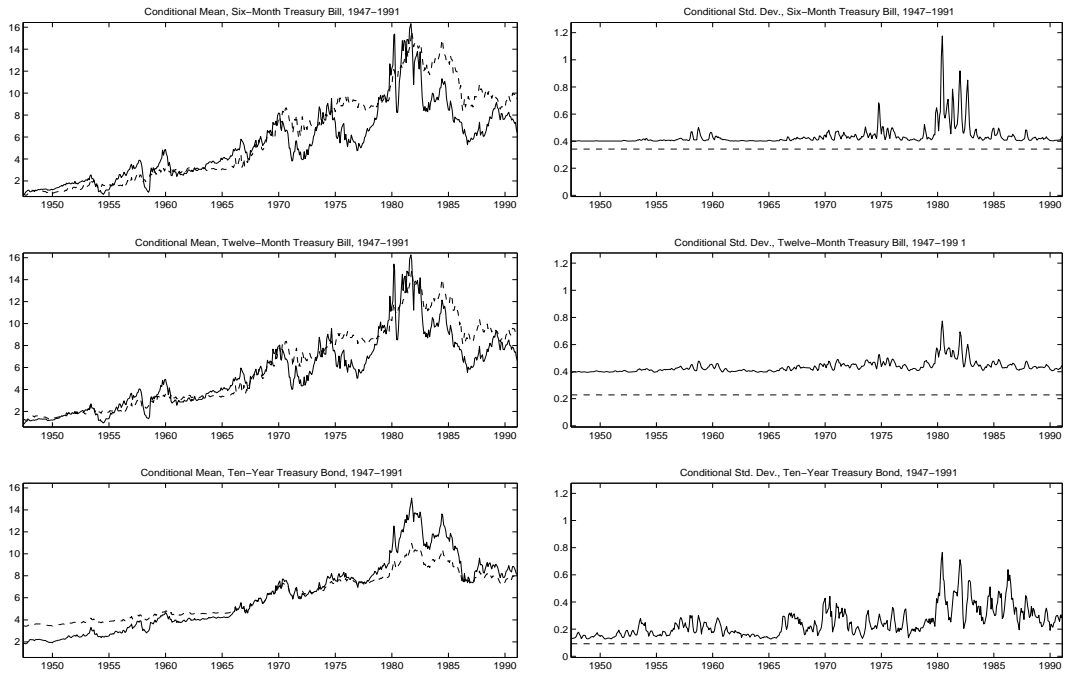Figure 3: Reprojected Conditional Moments: $A_0(1)Q(2)I(0)$

(a) Conditional Means   (b) Conditional Volatility

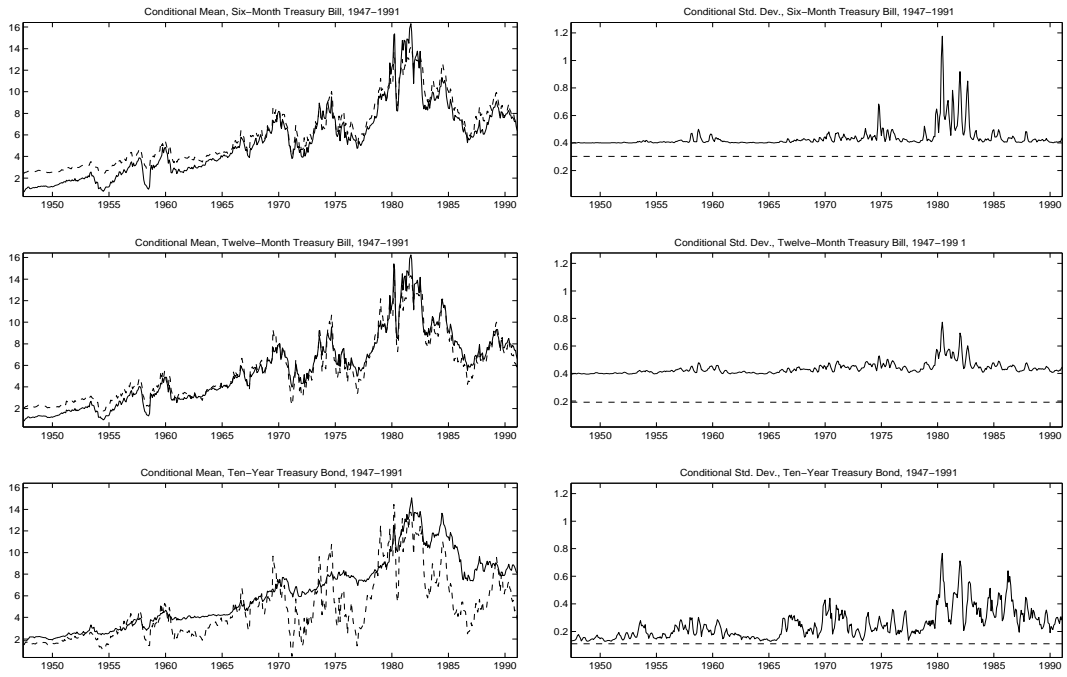Figure 4: Reprojected Conditional Moments: $A_2(2)Q(1)I(0)$

(a) Conditional Means   (b) Conditional Volatility

Figure 5: Reprojected Conditional Moments: $A_1(2)Q(1)I(0)$

(a) Conditional Means   (b) Conditional Volatility

Figure 6: Reprojected Conditional Moments: $A_2(3)Q(0)I(0)$