

Pricing Claims Under GARCH-Level Dependent Interest Rate Processes *

V. Cvsa[†] P. Ritchken[‡]

May 12, 2000

*The authors gratefully acknowledge the comments of seminar participants at the Fixed Income and Derivatives Conference at Yale School of Management, The Fields Institute in Toronto, the 1998 INFORMS meeting in Cincinnati, the RISK 2000 Congress in Paris, and the Interest Rate Models Conference sponsored by SIRIF in Edinburgh.

[†]Weatherhead School of Management, Case Western Reserve University, Cleveland, Ohio, 44106. Email vxc10@po.cwru.edu

[‡]Weatherhead School of Management, Case Western Reserve University, Cleveland, Ohio, 44106. Email phr@pyrite.som.cwru.edu.

Abstract

This article establishes a family of models for pricing interest rate sensitive claims when the underlying interest rate is driven by a two state variable GARCH process. Analytical solutions are established for the case when the innovations in the short rate are combinations of a normal and chi-squared random variables and the volatility of rates takes on a special GARCH form. GARCH models that nest level dependent interest rate models, including the Cox, Ingersoll, and Ross model are also considered. Algorithms are provided that permit the efficient pricing of American style interest rate claims under a rather broad array of dynamics, including GARCH and regime switching processes. The simple efficient algorithms for pricing interest rate claims that we establish should permit empiricists to use term structure and option data to more fully evaluate alternative volatility dynamics in interest rate markets.

Significant research has been conducted on the class of single factor affine yield models. The empirical evidence suggests that these models are too restrictive to fit nominal interest rate behavior.¹ As a result, researchers have considered more complex models. Examples include Longstaff and Schwartz (1992), who permit the short volatility to be stochastic, Ait-Sahalia (1996), who allows mean reversion and volatility to be related to the level of interest rates in rather complex ways, Gray (1996), who permits regime switching, and Brenner, Harjes and Kroner (1996), hereafter BHK, who incorporate GARCH effects into the dynamics of the spot rate. The intent with these models is to add additional realism primarily through the volatility process. The recent empirical tests by Gray and BHK on the time series of interest rates highlight the restrictions of many of the common single factor models, and the importance of incorporating GARCH like features into the dynamics.

BHK show that models which parameterize volatility only as a function of the level of interest rates over-emphasize the sensitivity of volatility to levels, and are unable to capture the serial correlation in conditional variances. They also show that simple GARCH models fail to capture the relationship between volatility and the level of rates. Their models, which incorporate both GARCH and Level effects, characterize the volatility process better than either Level or GARCH models alone. They conclude that there exists a strong need to establish theoretical option pricing models for interest sensitive claims that are driven by underlying GARCH-Level dependent processes.

In this paper we develop such models. Our underlying interest rate is modeled by a two state variable GARCH process, the first variable capturing a mean reverting short term rate and a second state variable capturing volatility. Our models are capable of handling a wide array of dynamics for the volatility. In particular, its dynamics could depend on recent interest rate innovations, the level of interest rates and other known information. For the special case when the volatility process does not depend on the level of rates, the two state variable GARCH process leads to analytical solutions for bond prices and selected derivative instruments. In this case, one period rates have distributions that are combinations of normal and chi-squared distributions. When the chi-squared innovation is shut down, the model reduces to a GARCH extension of the discrete Vasicek model. In this model, rates are conditionally normal for the single period, but, due to the GARCH feature, over multiple periods, the rate is not normal, and its distribution can display significant kurtosis and skewness that could be useful in explaining the volatility smile. Actually, since our analytical model also permits chi squared innovations in the riskless rate we are able to depart even further from normality with distributions that hopefully capture features that render the model a more realistic description of the process.

Several other models exist which incorporate stochastic volatility for the short rate. For

¹For an excellent discussion of these models and a summary of the empirical results see Chapter 11 of Campbell, Lo and MacKinlay (1997).

example, Longstaff and Schwartz (1992) have a two factor model and Chen (1996) has a three factor model of the term structure where the state variables are the short rate, the long run average of the short rate, and the local volatility. These models, like our analytical models, belong to the Duffie-Kan (1996) family, meaning that bond yields are affine in the state variables. Unlike stochastic volatility models, however, our GARCH models have some advantages in that very standard maximum likelihood estimation procedures can be used to estimate the parameters of the process.

In extending beyond the Duffie-Kan class, we seldom find analytical solutions for bond prices and therefore have to resort to numerical procedures for their computations. Estimating the parameters of the process using time series of bond prices then becomes a computational burden due to the necessity of repeated calls to the numerical routines.² In this paper we provide an efficient algorithm for computing interest rate claims, and, for models that are not in the Duffie-Kan family, we demonstrate how parameter estimation is feasible. For example, we can price American claims under a Cox, Ingersoll and Ross type model, extended to incorporate GARCH effects.

While our focus of attention is on developing model for pricing claims under GARCH, such models are not the only promising avenue for model development. Recently, several authors have begun studying the behavior of interest rates when the underlying dynamics follows a regime switching process. An early example of a regime switching model was established by Hamilton (1989). More recently, Gray (1996) has studied the time series properties of interest rates when the underlying regimes are characterized by different mean reversion and volatility updating schemes. Our results for option pricing under GARCH models can be generalized to regime switching models where the volatility updates depend not only on the current levels of rates and volatilities, but also on a second independent stochastic process. This extension permits option prices to be derived under processes like Hamilton or Gray.

The paper proceeds as follows. In the first section we specify the dynamics of a state variable and the pricing kernel in such a way that we obtain a model for interest rates that nests a large class of existing models and provides generalizations that may better explain derivative prices. We also identify the risk neutral measure under which all contingent claims can be readily priced. In section 2, we extend the Vasicek (1977) model to cases where the state variable has

²There are a few exceptions. For example, multifactor models in the Heath Jarrow Morton (1992) family, that are not in the affine class, have been investigated by Inui and Kijima (1998). These models are multivariate generalizations of the models developed by Ritchken and Sankarasubramanian (1995). Since these models have finite state variables, and analytical representations of the term structure, implementation of these models using Monte Carlo simulation is possible. While Monte Carlo simulation can be used to price European claims, the pricing of American claims is computationally expensive. This is true, despite recent remarkable advances in this area. For example, see the excellent review by Boyle, Broadie and Glasserman (1997), and Broadie and Glasserman (1998).

both normal and chi-square innovations and the volatility follows a GARCH like process. For this process, simple analytical equations exist for bond prices, and derivative contracts can be easily computed. In section 3 we show how to price American claims for a broad family of processes that are not necessarily in the Duffie Kan family. We illustrate the convergence rate of prices, and, for models not in the Duffie-Kan family, we demonstrate how a time series of short term liquid claims can be used to estimate the parameters of the process. In sections 4 and 5 we discuss the modifications of the algorithm that accommodate regime switching and the incorporation of chi-squared innovations in the state variable. The final section illustrates how the model can be made to fit an arbitrary initial term structure, and how some of the models can naturally be extended to handle multiple factors.

I The Basic Setup

In this section we establish an arbitrage free setting for pricing bonds and interest rate claims in a discrete time setting, when interest rate volatility follows a GARCH like process. To achieve this objective we need to model the stochastic discount factor, commonly referred to as the pricing kernel. Towards this goal, let m_t be the marginal utility of consumption of the representative investor. Following Campbell, Lo and MacKinlay (1997), we assume that the dynamics of this pricing kernel, m_{t+1}/m_t , is given by:

$$-\ln \frac{m_{t+1}}{m_t} = \delta_t + x_t + \lambda_t \epsilon_{t+1} + \phi_t \eta_{t+1} \quad (1)$$

where ϵ_{t+1} and η_{t+1} are independent standard normal random variables, and δ_t and λ_t and ϕ_t are predictable processes, the latter two representing the market prices of risk related to the two sources of uncertainty. The single period bond price is given by:

$$e^{-r_t} = E^P \left[\frac{m_{t+1}}{m_t} \middle| \mathcal{F}_t \right] = e^{-x_t - \delta_t + \lambda_t^2/2 + \phi_t^2/2}$$

where the expectation is taken under the probability generating measure, P . If we choose $\delta_t = \frac{\lambda_t^2}{2} + \frac{\phi_t^2}{2}$ then r_t inherits the properties of x_t .

Given a specification for the dynamics of the pricing kernel and the state variable, all the information that is necessary for pricing contingent claims is provided. While pricing of all claims can proceed, it is advantageous to construct a measure under which interest rate claims can be readily priced as if the local expectations hypothesis holds. Specifically, this change of measure reduces the number of state variables that are necessary for implementing algorithms for pricing interest rate claims.

Lemma 1

Assume the dynamics of the pricing kernel evolves as

$$-\ln \frac{m_{t+1}}{m_t} = \delta_t + x_t + \lambda_t \epsilon_{t+1} + \phi_t \eta_{t+1}.$$

Assume date T is the terminal date that we are considering and define measure Q by

$$\frac{dQ}{dP} = \frac{m_T}{m_0} e^{\sum_{i=0}^{T-1} r_i}.$$

Under measure Q , if Z_{t+1} is any integrable \mathcal{F}_{t+1} measurable random variable, then:

$$e^{-r_t} E^Q[Z_{t+1} | \mathcal{F}_t] = E^P[Z_{t+1} \frac{m_{t+1}}{m_t} | \mathcal{F}_t].$$

Under measure Q , $\epsilon_{t+1} \sim N(-\lambda_t, 1)$, and $\eta_{t+1} \sim N(-\phi_t, 1)$.

Proof:

See Appendix 1

We consider models where the dynamics of the state variable follows the process,

$$r_{t+1} = \alpha + \beta r_t + h_t r_t^{\gamma_1} \epsilon_{t+1} + \sigma r_t^{\gamma_2} (\epsilon_{t+1} - c)^2 \quad (2)$$

$$h_{t+1}^2 = F(h_t^2, r_t, \epsilon_{t+1}, \eta_{t+1}) \quad (3)$$

where $F()$ produces positive values. While, the algorithms we develop can handle rather general specifications for the updating process, $F()$, we shall focus on models where the update for the variable, h_{t+1}^2 , does not depend on the level of interest rates. The reason for this is that the first diffusion coefficient for the interest rate is $h_t \times r_t^{\gamma_1}$. Hence, the elasticity parameter, γ_1 , already captures the impact of interest rate level dependence on volatility.

An important family of models for the interest rate process that we consider is given by:

$$r_{t+1} = \alpha + \beta r_t + h_t r_t^{\gamma_1} \epsilon_{t+1} + \sigma r_t^{\gamma_2} (\epsilon_{t+1} - c)^2 \quad (4)$$

$$h_{t+1}^2 = a_0 + a_1 h_t^2 + a_{2,t} (\epsilon_{t+1} - d_t)^2 + a_{3,t} (\eta_{t+1} - e_t)^2, \quad (5)$$

where $a_{2,t}$, $a_{3,t}$, d_t and e_t are predictable processes. Under the risk neutral measure, Q , we then obtain:

$$r_{t+1} = \alpha + \beta r_t + h_t r_t^{\gamma_1} (\epsilon_{t+1}^* - \lambda_t) + \sigma r_t^{\gamma_2} (\epsilon_{t+1}^* - (c + \lambda_t))^2 \quad (6)$$

$$h_{t+1}^2 = a_0 + a_1 h_t^2 + a_{2,t} (\epsilon_{t+1}^* - (\lambda_t + d_t))^2 + a_{3,t} (\eta_{t+1}^* - (\phi_t + e_t))^2 \quad (7)$$

The dynamics under the data generating process, captured by equations (4) to (5), and the dynamics under the risk neutral process, given by equations (6) to (7), nest many well known models that can be categorized into three families.

The first family arises when $\sigma = 0$, and $a_{3,t} = 0$. Then we have:

$$\begin{aligned} r_{t+1} &= \alpha + \beta r_t + h_t r^{\gamma_1} \epsilon_{t+1} \\ h_{t+1}^2 &= a_0 + a_1 h_t^2 + a_{2,t} (\epsilon_{t+1} - d_t)^2. \end{aligned}$$

Here $\{h_t^2 | t = 0, 1, 2, \dots\}$ can be viewed as a rescaling process for the conditional variance of the interest rate, r_{t+1} . All the models considered by Chan Karolyi, Longstaff and Sanders (1992) are nested here, ($a_1 = a_{2,t} = 0$), as are the simple GARCH(1,1) models of interest rates, reviewed by Bollerslev, Chou, and Kroner (1992). ($\gamma_1 = 0$ and $a_{2,t} = a_2 h_t^2$). This model also includes Heston and Nandi (1999), who provide a model where GARCH effects are induced by normalized innovations ($\gamma_1 = 0$, $a_{2,t} = a_2$, and $d_t = dh_t$). Using time series data, BHK show that the effects of good and bad news on the volatility of interest rates is asymmetric. Specifically, negative shocks have a larger impact on volatility than the equivalent positive shocks. This suggests that $d_t > 0$.

The second family of models allow for chi-squared as well as normal innovations in the interest rate. In these models $a_{3,t} = 0$ in equation (5). Hence the rescaling factor, h_{t+1}^2 , has GARCH properties in that variances depend solely on the path of interest rate innovations.³ If, in addition, we set $a_0 = a_1 = a_{2,t} = 0$ in equation (5) and permit $\sigma > 0$ then innovations are chi squared, and we get models along the lines of Duan (1996b). Indeed, with $\gamma_2 = 1$, our model reduces to Duan. In general, we can also permit innovations to be combinations of normal and chi-squared random variables. The simplest example is a Vasicek like model, generalized to permit chi squared innovations as:

$$r_{t+1} = \alpha + \beta r_t + \sigma_1 \epsilon_{t+1} + \sigma_2 \epsilon_{t+1}^2.$$

which has no GARCH effects. It is well known that the Vasicek model does not produce yield curves that have sufficient curvature. Models that use both normal and chi-squared innovations may more readily capture the typical curvatures observed in the yield curve.⁴

A third family of models arise when $a_{3,t} > 0$. In this case the volatility updates do not only depend on interest rate innovations, but also on a second stochastic factor. As a result, these models generalize the typical GARCH models where variances depend solely on the path of interest rate innovations. In the extreme case, where $a_{2,t} = 0$ in equation (5), the rescaling factor does not depend on the interest rate innovation, and the model reduces to a regime switching process, where the regimes are directed by the stochastic process $\{\eta_t | t = 1, 2, \dots\}$. For example, a simple Hamilton (1989) type of regime switching process can be established. For example, in

³Constantinides (1992) has a model in which the logarithm of the pricing kernel has innovations which are the sum of non central chi squared random variables. Our approach is quite different, since our state variable, not the pricing kernel, has chi squared innovations.

⁴For a discussion on the poor performance of Vasicek models see Backus, Foresi, and Telmer (1997).

a two volatility regime switching model we have:

$$\begin{aligned}
r_{t+1} &= \alpha + \beta r_t + h_t r_t^{\gamma_1} \epsilon_{t+1} \\
h_{t+1}^2 | h_t^2 = \delta_1^2 &= \begin{cases} \delta_1^2 & \text{if } \eta_{t+1}^2 < k_1 \\ \delta_2^2 & \text{otherwise} \end{cases} \\
h_{t+1}^2 | h_t^2 = \delta_2^2 &= \begin{cases} \delta_2^2 & \text{if } \eta_{t+1}^2 > k_2 \\ \delta_1^2 & \text{otherwise} \end{cases}
\end{aligned}$$

For the case where $\gamma_1 = 0$, and conditional on the scaling factor, one period rates are normally distributed. Due to the random mixing of normal distributions, the tails of the distribution of rates over multiple periods will display fat tails.⁵

Before investigating algorithms for pricing claims under rather general GARCH volatility updating schemes, we consider cases that permit analytical solutions.

II Affine Yield Interest Rate Models

Consider the following special case of equations (4) - (5), under a pricing kernel given by equation (1), with $\lambda_t = \lambda h_t$. That is, the market price of risk is directly proportional to the rescaling factor, and under the data generating measure we have:

$$r_{t+1} = \alpha + \beta r_t + h_t \epsilon_{t+1} + \sigma \epsilon_{t+1}^2 \quad (8)$$

$$h_{t+1}^2 = a_0 + a_1 h_t^2 + a_2 (\epsilon_{t+1} - dh_t) + a_3 (\eta_{t+1} - eh_t)^2 \quad (9)$$

Let $P(t, t+n)$ represent the date t price of a discount bond that pays \$1 at date $t+n$.

Proposition 1

Assume the dynamics of the interest rate are given by the above equations. Further, assume ϕ_t and λ_t are proportional to h_t . Then, bond prices are given by the following recursive equation.

$$P(t, t+n) = e^{-A_n - B_n r_t + C_n h_t^2} \quad \text{for } n \geq 1. \quad (10)$$

where $A_1 = C_1 = 0$ and $B_1 = 1$, with

$$\begin{aligned}
A_{n+1} &= A_n + \alpha B_n - a_0 C_n + \frac{1}{2} \log(1 - 2(a_2 C_n - \sigma B_n)) + \frac{1}{2} \log(1 - 2a_3 C_n) \\
B_{n+1} &= (1 + \beta B_n) \\
C_{n+1} &= (\lambda - \sigma \lambda^2) B_n + a_1 C_n + (\lambda + d)^2 a_2 C_n + \\
&\quad \frac{(B_n - 2\lambda \sigma B_n + 2a_2 C_n (\lambda + d)^2)}{2(1 - 2(a_2 C_n - \sigma B_n))} + \frac{a_3 (e + \phi)^2 C_n}{1 - 2a_3 C_n}
\end{aligned}$$

⁵For further discussions on pure regime switching models see Hamilton (1989).

Proof:

See Appendix 1

Note that when $a_1 = a_2 = a_3 = \sigma = 0$, the model reduces to the Vasicek model, discussed by Backus, Foresi and Telmer (1997). When the chi-squared innovations are switched off, ie $\sigma = 0$, as well as $a_3 = 0$, the model reduces to a Vasicek type model, extended to allow for GARCH effects. In this case, while single period conditional yields are normal, yields over multiple periods will not be normal. For example, the kurtosis of the distribution is largely determined by the leverage parameter, d . In addition, the model has two state variables, that are correlated, rather than the single state variable of the Vasicek model. Figure 1 illustrates how kurtosis and skewness in the distribution of interest rates can easily be controlled by the leverage parameter, d .

[Insert Figure 1 Here]

The most general form of the model, with $a_3 = 0$, allows interest rate innovations to be combinations of normal and dependent chi-squared random variables. These innovations, together with the GARCH effects make the model quite distinct from the Vasicek model, and has the potential to remove the well known volatility smile biases that exist when normal or lognormal processes are used in pricing interest rate claims.

Finally, when $a_3 > 0$, the variance process depends not only on the path of interest rates, but also on the stochastic process for η_t . A very simple case of this model occurs when the GARCH effects are shut down ($a_1 = a_2 = 0$) and $\sigma = 0$. In this case, the variance process is fully determined by the orthogonal state variable, and since interest rate over multiple periods are mixtures of normals, their distributions will display fat tails.

Proposition 1 easily extends to multi-factor versions. As an example, in Appendix 2, we develop analytical solutions for a specific two stochastic driver model, which contain up to four state variables.

Heston and Nandi (1999) have independently derived the analytical solution for the special case of this model when $a_3 = \sigma = 0$. Following along the lines of Foster and Nelson (1994), they show that a continuous time limit of this model can be represented as:

$$\begin{aligned} dr(t) &= (\alpha + \beta r(t))dt + \sigma(t)dw_1(t) \\ d\sigma^2(t) &= \square(\theta - \sigma^2(t))dt + k\sigma(t)dw_1(t) \end{aligned}$$

Since this continuous time model has drifts and volatilities that are affine in $r(t)$ and $\sigma^2(t)$, it belongs to the Duffie-Kan family, and analytical solutions are available for bond prices in the diffusion economy. The advantage of the discrete time formulations rests primarily on parameter

estimation. Generally, estimating the parameters of GARCH processes are trivial relative to the estimation issues associated with continuous time diffusions.

If the underlying interest rate has dynamics different from that in Proposition 1 then analytical solutions for bond prices may not be available. For example, consider the interest rate dynamics given by

$$r_{t+1} = \alpha + \beta r_t + h_t \epsilon_{t+1} \quad (11)$$

$$h_{t+1}^2 = a_0 + a_1 h_t^2 + a_2 h_t^2 (\epsilon_{t+1} - \lambda)^2. \quad (12)$$

In this model the scaling factor updates depend on the *total* innovation in the interest rate, rather than on the *normalized* innovation, used in Proposition 1. Following Duan (1997), it can be show that this type of process has a limiting diffusion representation of the form:

$$\begin{aligned} dr(t) &= (\alpha^* + \beta^* r_t) dt + \sigma(t) dw_1(t) \\ d\sigma^2(t) &= a_1^* dt + b_1^* \sigma^2(t) dw_1(t) + b_2^* \sigma^2(t) dw_2(t) \end{aligned}$$

where $E(dw_1(t)dw_2(t)) = 0$.

Notice that the discrete representation for interest rates by a two state variable model has just one stochastic driver while the continuous diffusion limit can be represented by two stochastic drivers. The diffusion limit does not belong to the Duffie Kan class, and no simple analytical representations for bond prices exist.

The models identified in Proposition 1 have the advantage that the time series of different bond prices can be used to estimate the parameters. In addition, since the entire yield curve can be constructed once the state variables are known, efficient numerical schemes can be established to price complex interest rate claims. However, since these models fall in the Duffie Kan family, they may not be flexible enough to capture important dynamics in the yield curve. As a result, it may be advantageous to consider algorithms for pricing claims under interest rate dynamics that differ from those in Proposition 1.

III Lattice Models for Pricing Interest Rate Claims

The model that we develop is related to the GARCH option pricing model for *equity* claims developed by Ritchken and Trevor (1999) (hereafter RT). Unfortunately, adopting their model to price interest rate claims is not viable for several reasons. First, in order to price interest rate claims, it may be necessary to have available the entire term structure of yields (ie. a vector) at each setting of the two state variables. In the RT algorithm, the number of distinct settings of the state variables is enormous, since only one of the two state variables, namely the logarithm of the stock price, has its domain restricted to a grid of values. As a result, the number of term

structures that need to be built up successively, using backward recursion, grows exponentially with the number of time periods. Our first modification of the RT algorithm is to discretize the volatility levels so that only a finite number of yield curves need to be built up, regardless of the number of time periods in the lattice. Second, with interest rate dynamics, the effect of mean reversion can be significant and its effects need to be incorporated into the design of an algorithm. We therefore modify the RT algorithm in how probability values are assigned to successor state variables. Third, in RT, the second state variable was the volatility. In our case, the second state variable is a scaling factor, and the volatility of rates could be a complex function of this variable. This creates problems when establishing the probabilities of moves in the state variables, from one set to the next. The algorithm that we establish sets up a stationary Markov chain approximation to the dynamics, in which the successor state variables and their transition probabilities are clearly identified in a first phase or forward scan. This approach is quite different to that in Ritchken and Trevor. Finally, in the RT algorithm, only GARCH processes were considered. In our case we permit a second uncorrelated random variable to affect our updating scheme.

We first consider pricing interest rate claims when the underlying rates follow a *finite discretized GARCH process* with conditional normally distributed innovations. Specifically, under the data generating measure, such a process is defined by:

$$r_{t+1} = \alpha + \beta r_t + h_t r_t^\gamma \epsilon_{t+1} \quad (13)$$

$$h_{next}^2 = a_0 + a_1 h_t^2 + a_2 h_t^2 (\epsilon_{t+1} - d)^2$$

$$h_{t+1}^2 = \delta_j^2 \text{ if } \frac{\delta_{j-1}^2 + \delta_j^2}{2} < h_{next}^2 < \frac{\delta_j^2 + \delta_{j+1}^2}{2} \text{ for } j = 1, 2, \dots, K \quad (14)$$

where, we define $\delta_0 = -\delta_1$ and $\delta_{K+1} = \infty$.

For pricing purposes, under the Q -measure, we have:

$$r_{t+1} = \alpha + \beta r_t + h_t r_t^\gamma (\epsilon_{t+1} - \lambda_t) \quad (15)$$

$$h_{next}^2 = a_0 + a_1 h_t^2 + a_2 h_t^2 (\epsilon_{t+1} - (d + \lambda_t))^2$$

$$h_{t+1}^2 = \delta_j^2 \text{ if } \frac{\delta_{j-1}^2 + \delta_j^2}{2} < h_{next}^2 < \frac{\delta_j^2 + \delta_{j+1}^2}{2} \text{ for } j = 1, 2, \dots, K \quad (16)$$

When $\gamma = 0$, the above model only permits the local volatility to take on K distinct values. However, with $\gamma_1 > 0$, local volatilities are not curtailed to K values. While such models may be of some interest in their own right, in our case they are particularly important, since, for this family, we can develop efficient algorithms for pricing interest rate claims. Moreover, as we shall see, these type of models converge to GARCH models as the partitioning scheme is refined.

We begin by approximating the sequence of single period conditionally normal random variables for the interest rate in equation (15) by a sequence of discrete random variables that can

take on values on a grid. Let Δ be a fixed constant that determines the gap between adjacent interest rates on a grid, and let r_0 be the given initial interest rate. The interest rates in subsequent periods are restricted to values $r_0 + j\Delta_m$ where $j = 0, \pm 1, \pm 2, \pm 3, \dots$ and $\Delta_m = \Delta/\sqrt{m}$. Here the subscript m denotes values derived from using a $2m + 1$ point approximation to the conditionally normal random variable. The idea is that as m increases, the approximation improves.

Assume the state variables at date t are (r_t, δ_k^2) , where r_t is on the grid of values and $k \in \{1, 2, 3, \dots, K\}$. Then, the random variable, r_{t+1} , is a normal random variable with mean \bar{E}_t and variance δ_k^2 , where:

$$\bar{E}_t = \alpha + \beta r_t - \lambda_t \delta_k r_t^\gamma$$

Of course, there is no reason that \bar{E}_t will be a grid point. Let ξ_1 be an integer, such that $r_t + \xi_1 \Delta_m$ is the nearest grid point to \bar{E}_t . To approximate the conditional normal random variable, we restrict the movements of r_{t+1} to $2m + 1$ values, of which m values are above and below $r_t + \xi_1 \Delta_m$, with the middle value equal to $r_t + \xi_1 \Delta_m$.

The exact grid points that are used for the approximation are tentatively determined by putting $\xi_2 = 1$ and using the points

$$r_t + \xi_1 \Delta_m \pm j \xi_2 \Delta_m, \text{ for } j = 0, \pm 1, \pm 2, \dots, \pm m$$

If the conditional variance of the interest rate at the “node” (r_t, δ_k) is “small” then discrete probability values can be found over the grid of $2m + 1$ interest rates that surround the expected interest rate, such that the conditional first two moments of the discrete variable matches the true values of the conditional normal distribution being approximated. In this case $\xi_2 = 1$ will suffice. However, if δ_k is sufficiently large, and the level of rates are high, then the local variance may be large and it may not be possible to find valid probability values for the surrounding $2m + 1$ interest rates such that the first two moments match. In this case, we construct an approximating distribution that uses every second interest rate point on the grid surrounding the expected value. If these “double-sized” jumps ($\xi_2 = 2$) still do not permit the means and variances to be exactly matched while producing valid probabilities, then the approximating scheme uses points separated by larger Δ_m multiples. In general, ξ_2 is the positive integer closest to zero that allows the mean and variance of next period’s interest rate to be matched to the true moments while at the same time ensuring that all the $2m + 1$ probability values are valid numbers in the interval $[0, 1]$. We refer to ξ_2 as the jump parameter. This idea, of stretching out the successor nodes on a grid of points was first used by Ritchken and Trevor (1999) in pricing equity options under GARCH specifications. In their approach, however, they assumed $\xi_1 = 0$. When interest rates are mean reverting it makes more sense to shift the approximation so that the middle jump is centered close to the expected value.

Let $P(\xi_1 + j\xi_2 | r_t, \delta_k)$ represent the probability that the interest rate moves from r_t to $r_t + (\xi_1 + j)\Delta_m$. for $j = 0, \pm 1, \pm 2, \dots, \pm m$. These probabilities are chosen such that:

$$P(\xi_1 + j\xi_2 | r_t, \delta_k) = \sum_{j_u, j_m, j_d} \binom{n}{j_u \ j_m \ j_d} p_u^{j_u} p_m^{j_m} p_d^{j_d} \quad (17)$$

with $j_u, j_m, j_d \geq 0$ such that $n = j_u + j_m + j_d$ and $j = j_u - j_d$. The expression in brackets denotes the trinomial coefficient $\frac{n!}{j_u! j_m! j_d!}$ and the trinomial probability values p_u, p_m and p_d are values between zero and one that satisfy:

$$\begin{aligned} p_u + p_m + p_d &= 1 \\ (\xi_1 \Delta_m / m + \xi_2 \Delta_m) p_u + (\xi_1 \Delta_m / m) p_m + (\xi_1 \Delta_m / m - \xi_2 \Delta_m) p_d &= E_t^Q[\Delta r_t] \\ (\xi_1 \Delta_m / m + \xi_2 \Delta_m)^2 p_u + (\xi_1 \Delta_m / m)^2 p_m + (\xi_1 \Delta_m / m - \xi_2 \Delta_m)^2 p_d &= E_t^Q[\Delta r_t^2] \end{aligned}$$

where Δr_t is a normal random variable representing the change in interest rates over a period from $[t, t + \frac{1}{m}]$ given the mean is \bar{E}/m and the variance is $(\delta_k^2 r_t^{2\gamma})/m$.⁶

For each interest rate on the grid, and for each rescaling factor, these probability values, together with ξ_1 and ξ_2 need to be computed. The interest rate innovation term, ϵ_{t+1} is then approximated by ϵ_{t+1}^a , where

$$\epsilon_{t+1}^a = \frac{j\xi_2 \Delta_m + (\xi_1 \Delta_m) - (\alpha + (\beta - 1)r_t - \lambda_t \delta_k r_t^\gamma)}{\delta_k [r_t]^\gamma} \quad (18)$$

and $j = 0, \pm 1, \pm 2, \dots, \pm m$. Viewed from time t , ϵ_{t+1}^a is a discrete state random variable, with mean 0 and variance 1, that converges in distribution to a continuous state standard normal random variable as $m \rightarrow \infty$.

Given, the interest rate innovation, ϵ_{t+1}^a , the statistic h_{next}^2 can be computed, and the next scaling factor can be identified. In particular, given the state variables are (r_t, δ_k^2) , for pricing purposes, we have:

$$r_{t+1} = r_t + \xi_1 \Delta_m + j\xi_2 \Delta_m \quad (19)$$

$$\begin{aligned} h_{next}^2 &= a_0 + a_1 \delta_k^2 + a_2 \delta_k^2 (\epsilon_{t+1}^a - (d + \lambda_t))^2 \\ h_{t+1}^2 &= \delta_j^2 \text{ if } \frac{\delta_{j-1}^2 + \delta_j^2}{2} \quad h_{next}^2 < \frac{\delta_j^2 + \delta_{j+1}^2}{2} \text{ for } j = 1, 2, \dots, K. \end{aligned} \quad (20)$$

In summary, then, for each interest rate on the grid, and for each of the K rescaling levels, we have computed, ξ_1, ξ_2 , the $2m + 1$ probability values, and the $2m + 1$ successor rescaling levels. This completes the specification for the approximating process.

⁶To approximate the normal random variable over a single period by a discrete $(2m+1)$ -nomial random variable, first partition the time period into m sub-intervals. Over each subinterval apply a trinomial approximation to the normal random variable over the increment. The three equations above ensure that the trinomial distribution has the same first two moments as the normal distribution. Here, p_u, p_m and p_d represent the probability of an ‘‘up’’, ‘‘middle’’ and ‘‘down’’ jump, where ‘‘middle’’ is defined according to the drift, or more precisely by ξ_1 . Given the sequence of m independent trinomials, the total probability to each terminal ‘‘node’’ is given by equation (17).

Proposition 2

Let Δ be a given constant and $\Delta_m = \Delta/\sqrt{m}$. Define r_0 to be the initial interest rate, and restrict its future values to the points $r_0 + j\Delta_m$ for $j = 0 \pm 1, \pm 2, \dots$. Let $\{\delta_1, \delta_2, \dots, \delta_K\}$ be the set of scaling values, with $h_0 = \delta_k$ given, for some k , $k = 1, 2, \dots, K$. The Markov chain $\{(r_t, h_t) \mid t = 0, 1, 2, \dots\}$ characterized by equations (19) to (20), converges in distribution to a discrete GARCH regime process $\{(r_t, h_t) \mid t = 0, 1, 2, \dots\}$ given in equations (15)- (16) as $m \rightarrow \infty$.

Proof:

See Appendix 3.

Pricing interest rate claims on the lattice is now rather straightforward. Assume that there are a total of R different interest rate levels on the lattice and K scaling factor levels (regimes). Then, we will essentially have $R \times K$ distinct term structures. These term structures can be built up sequentially using backward recursion on the lattice and by pricing bonds under the local expectations hypothesis in the Q measure. Once all the term structures have been built up, then interest rate claims can be readily priced. Specifically,

$$\begin{aligned} C(r_t, \delta_k) &= e^{-r_t} E_t^Q [C(r_{t+1}, h_{t+1} \mid r_t, \delta_k)] \\ &= e^{-r_t} \sum_{j=-m}^m C(r_t + \xi_1 \Delta_m + j \xi_2 \Delta_m, \delta_{k_j}) P(\xi_1 + j \xi_2 \mid r_t, \delta_k) \end{aligned}$$

In the above equation δ_{k_j} represents the scaling factor that follows level k given that the interest rate is r_t , the current scaling factor is δ_k , and the current innovation is a jump of size $(\xi_1 + j \xi_2) \Delta_m$. Notice that the ξ_1 and ξ_2 values will vary depending on the level of the state variables. If the claim is an American claim, then the value, $C(r_t, \delta_k)$, has to be compared with the intrinsic value of the claim given that it is exercised.

While the finite GARCH regime model is of some interest in its own right, as discussed earlier, the model really serves as an important bridge to continuous state GARCH - Level dependent interest rate processes, and hence to stochastic volatility models. Again, to make the exposition of this convergence clear we continue to consider our model given in equations (19) and (20).

Now, assume that the partition of the state space for the second state variable, i.e. for the rescaling levels, is set up such that as K increases, the mesh of the partition decreases. In particular,

1. As $K \rightarrow \infty$, $\delta_K^2 \rightarrow \infty$ and $\delta_1^2 \rightarrow a_0$
2. $Max_{j \neq n} |\delta_j^2 - \delta_{j-1}^2| \rightarrow 0$ as $K \rightarrow \infty$.

This leads to:

Proposition 3

Under the above two conditions, the discretized GARCH process converges pointwise to the continuous GARCH process given below.

$$\begin{aligned}r_{t+1} &= \alpha + \beta r_t + h_t r_t^\gamma (\epsilon_{t+1} - \lambda_t) \\h_{t+1}^2 &= a_0 + a_1 h_t^2 + a_2 h_t^2 (\epsilon_{t+1} - (d + \lambda_t))^2\end{aligned}$$

Proof:

See Appendix 3.

Taken together, Propositions 2 and 3 imply that a Markov Chain $\{(r_t, h_t) \mid t = 0, 1, 2, \dots\}$ can be established such that as m and $K \rightarrow \infty$ the discretized GARCH process converges in distribution to the above continuous state GARCH process.

To illustrate the rate of convergence of the finite GARCH regime option prices to their continuous state GARCH level dependent limits, we first consider the GARCH-Vasicek model, for which analytical solutions for the yield curve are available. In our implementation, we fixed the number of periods at 120 months, and we approximated the number of rescaling factors by K values where K increases from 1 to 30 equally spaced values. The spacing is determined such that a_0 is the lowest variance state, and the initial volatility, h_0 , is the middle state. All the algorithms for the models were run on a Pentium 233 MHz processor and the times (in seconds) are reported in the tables.

Table 1 shows the convergence of selected yields to their theoretical values for the pure Vasicek model, where there is only one volatility level. The results obtained using a trinomial approximation over each month are compared to the analytical solutions. As can be seen the trinomial approximation is effective.

Table 2 shows the convergence of prices to the GARCH like Vasicek model as the number of rescaling levels increase. The first row shows that the pure Vasicek model, with one rescaling factor, provides a poor approximation. However, as the number of rescaling levels increase the yields begin to stabilize. Prices of an at-the-money one year option on a two year bond are shown in Figure 2. Like the bond prices, the convergence rate of prices as the number of scaling factors increase is fairly rapid.⁷

[Insert Table 1 and 2 Here]

[Insert Figure 2 Here]

Table 3 repeats the analysis of Table 1 for the pure CIR model. As with the pure Vasicek model, the algorithm performs well, even for the case $m = 1$. Table 4 shows the results for the

⁷The tables and figures shown here represent very typical convergence patterns.

GARCH-CIR model. While there are no analytical solutions for yields or options, the numerical results indicate that yields stabilize quite rapidly, and that there is little benefit in partitioning the rescaling factors with more than 40 states.

[Insert Tables 3 and 4 Here]

Figure 3 shows the typical convergence rate of option prices in the GARCH-CIR model. As the number of scaling factors increase to 40 the option price converges.

[Insert Figure 3 Here]

As discussed earlier, the cost of moving beyond the affine family could be high since parameter estimation, using the time series of bond and perhaps other liquid derivative prices, requires repeated calls of the numerical routines. Several different approaches can be used to simplify the estimation problem. For example, historical interest rates can be used to estimate all the parameters under the data generating measure. The additional parameters, namely the market prices of risk, can be extracted from the time series of prices of selected liquid instruments. Moreover, if we restrict the cross section of claims to the most liquid short term instruments, then the numerical procedures will be reasonably efficient.

To illustrate this, we consider the time it takes to compute a single iteration in which 4 short term discount bonds and 4 short term caplets are priced for a given parameter set. Specifically, a 3, 6, 12, and 24 month discount bond is priced along with 3 month caplets with expiration dates in 3, 6 and 12 months. The computational times are somewhat insensitive to the parameter settings, but the specific parameter values were selected to be consistent with the values obtained in empirical studies by Backus, Foresi and Telmer (1997). The set of prices takes less than 5 seconds to compute on a Pentium 233 MHz processor. The total computational time for a given parameter set will depend on how many cross sectional prices are used in the analysis. If the time series consists of n sets of these prices then n sets of theoretical prices have to be obtained. This will take at most $5 \times n$ seconds.⁸ The total computational effort for the optimization problem will depend on the number of calls required to identify an optimal solution. This typically depends on the number of free parameter values and upon the initial solution.⁹

⁸Actually, if n is large, then, for a given set of parameters, the computational time can be reduced dramatically from this bound by solving one larger problem over a much wider range of initial state variables. The prices of the short dated claims are then obtained over a large fine grid of the state variables. For any realization of the state variables over time, the prices of the short dated instruments can be approximated by appropriate interpolation of prices. In our example, the gap size between successive interest rate levels was about 10 basis points. If this is reduced to about 2-4 basis points then 20-30 seconds of computational effort is sufficient to generate all pricing information, regardless of n .

⁹Our experience with say 3 unknown parameter values, is that the optimizations may require from 10 to 1000 calls, indicating a wide range of computational effort.

Once the parameters are estimated, then of course, less efficiency is required for the pricing of any specific American interest rate claim. If the lattice is to be used to price contracts when the underlying dynamics satisfy the conditions of Proposition 1, then the algorithm simplifies since the term structure is explicitly known for each setting of the state variables. For example, in pricing a three month option on a thirty year bond, a lattice need only be built out to three months. Similarly, if a Bermudan swaption is to be priced, then the value of early exercise at any intermediate setting of the two state variables, can easily be assessed. If no analytical solutions are available, then the exercise value at each exercise date has to be numerically computed.¹⁰

IV Regime Switching Models

The above lattice based model can be extended to include the effects of the orthogonal process generated by the sequence $\{\eta_t | t = 1, 2, \dots\}$. Consider the model:

$$\begin{aligned} r_{t+1} &= \alpha + \beta r_t + h_t r_t^\gamma \epsilon_{t+1} \\ h_{next}^2 &= a_0 + a_1 \delta_k^2 + a_2 \delta_k^2 (\epsilon_{t+1} - d)^2 + a_3 \delta_k^2 (\eta_{t+1} - e)^2 \\ h_{t+1}^2 &= \delta_j^2 \text{ if } \frac{\delta_{j-1}^2 + \delta_j^2}{2} < h_{next}^2 < \frac{\delta_j^2 + \delta_{j+1}^2}{2} \text{ for } j = 1, 2, \dots, K \end{aligned}$$

where, as before, we define $\delta_0 = -\delta_1$ and $\delta_{K+1} = \infty$.

For pricing purposes, under the Q -measure, we have:

$$\begin{aligned} r_{t+1} &= \alpha + \beta r_t + h_t r_t^\gamma (\epsilon_{t+1} - \lambda_t) \\ h_{next}^2 &= a_0 + a_1 \delta_k^2 + a_2 \delta_k^2 (\epsilon_{t+1} - (d + \lambda_t))^2 + a_3 \delta_k^2 (\eta_{t+1} - (e + \phi_t))^2 \\ h_{t+1}^2 &= \delta_j^2 \text{ if } \frac{\delta_{j-1}^2 + \delta_j^2}{2} < h_{next}^2 < \frac{\delta_j^2 + \delta_{j+1}^2}{2} \text{ for } j = 1, 2, \dots, K \end{aligned}$$

Pricing claims in such models can readily be accomplished with a small modification to the algorithm. In particular, the model is still a two state variable model. Given the state is (r_t, δ_k) , to price a claim under the risk neutral measure proceeds on the lattice as before, except that there is an additional step. Specifically, given an interest rate innovation, the value of h_{next}^2 is no longer certain, and hence the successor scaling factor is not uniquely established. However, its value will be uniquely determined once the value of η_{t+1} is known. Given a discretization of η_{t+1} , the conditional value of the scaling factor, $h_{next}^2 | \eta_{t+1}$, can be established, and the next scaling factor identified. Since, the random variables, η_{t+1} and ϵ_{t+1} are independent, the probability distribution of joint moves in the interest rate and orthogonal volatility innovations can be

¹⁰For example, given the parameters, the computational time for a 5 year Bermudan swaption with six month exercise dates should be under 15 seconds.

established, and the joint expected price of the claim, conditional on the state variables, can be computed. Specifically, we have

$$\begin{aligned} C(r_t, \delta_k) &= e^{-r_t} E_t^Q [C(r_{t+1}, h_{t+1} | r_t, \delta_k)] \\ &= e^{-r_t} \sum_{i=-n}^n \sum_{j=-m}^m C(r_t + \xi_1 \Delta_m + j \xi_2 \Delta_m, \delta_{s_j(i)} P(\xi_1 + j | r_t, \delta_k) \pi(i)) \end{aligned}$$

where $s_j(i)$ is the next scaling factor that is completely determined once the interest rate innovation, indexed by j , and orthogonal volatility innovation, indexed by i , is determined, and $\pi(i)$ is the probability that the approximating standard normal random variable falls in the i^{th} interval.¹¹

V Non Gaussian Innovations in the State Variable

The above algorithm can also be modified to handle chi squared innovations in the state variable. In developing the previous algorithm, we relied on the fact that we could obtain the innovation term, ϵ_{t+1}^a , given that we have information on r_t , r_{t+1} and h_t . Specifically, ξ_1 and ξ_2 were chosen such that the first two moments of the approximating process in equation (19) matched the true moments, while the probability values on the lattice are all between 0 and 1. Then, equation (18) was used to compute ϵ_{t+1}^a . This value is then used to compute the updated variance, h_{t+1}^2 , in equation (20). With the introduction of chi-squared innovation, it is not possible to recover the innovation term uniquely. Hence the algorithm has to be modified.

We first fix a grid of values for the state variable, r_t , with adjacent points separated by Δ , say. We also assume that there are a finite number, K , of rescaling factors. Assume that we are at a particular value for r_t and h_t that are on the grid of values. We approximate the normal innovation term ϵ_{t+1} using a $2m+1$ -nomial random variable. Let $P(j)$ represent the probability of the j^{th} jump, with $j = 0, \pm 1, \pm 2, \dots, \pm m$. By using each of these points, we can compute the $2m+1$ successor pairs of state variables, r_{next} and h_{next} . In particular, under the risk neutral process, we have

$$\begin{aligned} r_{next} &= \alpha + \beta r_t + h_t [r_t]^{\gamma_1} (\epsilon_{t+1}^a - \lambda_t) + \sigma (\epsilon_{t+1}^a - (c + \lambda_t))^2 \\ h_{t+1}^2 &= F(h_t^2, \epsilon_{t+1}^a, r_t). \end{aligned}$$

The backward recursion method for pricing claims proceeds as follows. Assume the current “node” is (r_t, δ_k) . We begin by establishing the next $2m+1$ interest rates, and the next rescaling factor. The updated value of the rescaling variable, δ_{k_j} say, is completely determined by the

¹¹Approximating the $2n+1$ point approximation for the standard normal random variable, η_{t+1} , can be done in the same way as for the approximating $2m+1$ point approximation for ϵ_{t+1} .

current value, δ_k , by the current interest rate, r_t , and by the interest rate innovation, which is itself fully determined by the j^{th} value of ϵ_{t+1}^a . Let r_{next_j} be the next approximating interest rate, based on the j^{th} jump of ϵ^a , where $j = 0 \pm 1, \dots, \pm m$. Note that the successor value, r_{next_j} , may not necessarily fall on the grid of feasible values at date $t + 1$. For each of these interest rates, we have to identify the two surrounding interest rates on the grid. Let r_{t+1,u_j} and r_{t+1,d_j} be the interest rate on the closest grid point above and below r_{next_j} . Let p_j ($1 - p_j$) be the interpolated weight assigned to the upper (lower) grid point. That is

$$p_j = 1 - \frac{r_{t+1,u_j} - r_{next_j}}{\Delta}$$

At each of these two points, $(r_{t+1,u_j}, \delta_{k_j})$ and $(r_{t+1,d_j}, \delta_{k_j})$ we pick up the prices of the claim, $C(r_{t+1,u_j}, \delta_{k_j})$ and $C(r_{t+1,d_j}, \delta_{k_j})$ and make an appropriate interpolation to establish the theoretical value. This value is discounted at r_t . This interpolation is repeated at all the $2m + 1$ successor interest rates. The final price of the claim is obtained by weighting these $2m + 1$ prices by their probabilities, $P(j)$, where $j = 0, \pm 1, \pm 2, \dots, \pm m$. We have:

$$C(r_t, \delta_k) = e^{-r_t} \sum_{j=-m}^m [p_j C(r_{t+1,u_j}, \delta_{k_j}) + (1 - p_j) C(r_{t+1,d_j}, \delta_{k_j})] P(j)$$

As before, if the claim is American, then this value has to be compared to the intrinsic value at this node. To obtain the intrinsic value, it may be necessary to first reconstruct the full term structure at each setting of the state variables, (r_t, δ_k) on the grid.

The interpolations has the effect of increasing the variance, in such a way that the first two moments on the lattice are a bit larger than their true moments. Convergence of prices on the lattice, however, is guaranteed as m (which determines the $2m + 1$ -nomial random variable used to approximate the innovation term ϵ_{t+1}) goes to ∞ and as Δ (which determines the spacing of the interest rates on the grid) goes to 0.

To illustrate the convergence behavior on the lattice we consider bond prices using our pure chi squared process, for which we have analytical solutions. Table 5 shows the convergence of selected yields and option prices to the theoretical values as the approximating process for the normal distribution is refined. The analytical solutions for the yields were obtained using the restricted version of the model in Proposition 1.

[Insert Table 5 Here]

VI Conclusion

This article has addressed the important problem of pricing interest rate claims under a large family of two state variable processes, driven by either one or two stochastic drivers. The models

that we establish are all discrete time models in which the volatility follows a GARCH and/ or a Regime Switching process. The models nest most of the single state variable models such as the discretized version of the Vasicek and Cox, Ingersoll, Ross models. In addition, the innovation terms were generalized to include not only Gaussian but chi-squared terms. The models therefore allow interest rates to take on distributions quite distinct from normal distributions. By allowing distributions to have skewness and kurtosis, the models have a good chance of removing the volatility smile that is present when normal and lognormal processes are used in pricing claims such as interest rate caps and swaptions. Further, since many of the GARCH processes have diffusion limits that lead to stochastic volatility models, these models can be viewed as useful approximations to their diffusion limit counterparts. The advantage of the GARCH models, is that estimating the parameters of the process is straightforward, whereas a direct estimation of a non observable volatility process is more complex.

This article presented analytical models for bonds when interest rate processes take on particular GARCH and Regime Switching forms. These models nest the Vasicek model, but since innovations need not be normal and since volatility follows a path dependent process, the distributions of rates are far more flexible than the Vasicek model. The benefits of such types of models in explaining the time series properties of interest rates have been well motivated by studies such as BHK. We now have provided models that permit interest rate claims to be priced. It remains for future empirical research to evaluate the impact of incorporating cross sectional derivative prices into the time series analysis.

While the analytical models presented here appear to be very flexible, additional generality can be established, at the expense of foregoing analytical tractability. A computational scheme has been established that permits interest rate claims to be priced in a discretized GARCH/Regime Switching model. The algorithm used to compute prices for a finite discretization of scaling factors has been shown to converge to prices of claims generated under continuous state GARCH processes when the partition of scaling factors is suitably refined.

There are two immediate applications of the models presented here. The first involves empirical studies to ascertain the volatility structures in interest rate markets. Since bond and other interest rate claims contain significant information on volatilities, it seems sensible to incorporate cross sectional data on these items into the time series analysis. Hence studies like Brown and Dybvig (1986) and Brown and Schafer (1994), in which term structure data was incorporated into the time series analysis to test Cox Ingersoll and Ross models can now be extended to incorporate other interesting models. As Campbell, Lo and MacKinlay note, it would be desirable to repeat the BHK study using cross sectional information on derivative security prices. The models developed here permit such studies to be performed.

The second immediate application is to price interest rate claims. Since our models contain a finite number of parameters, it is possible only to match a finite number of points on an initial

term structure. If traders desire models that match the term structure exactly, then this can easily be accommodated by first subtracting the theoretical forward rates, developed using a model calibrated to fit the finite number of points, from actual forward rates. This yields a set of residuals, which can be attributed to a *deterministic* term structure model. Specifically, the residuals are exactly equal to the logarithm of the pricing kernel of an independent (deterministic) term structure model. By multiplying the theoretical bond prices with the bond prices based off this second term structure model, one will recover the original term structure.¹² Current interest rate claim trading models heavily weight current information on the term structure and on volatilities in calibrating parameters. Our models allow traders to incorporate historical time series data into the analysis. It remains for future research to establish how heavily historical data should be weighted in such a process.

While the focus of this paper has been on two state variable models driven by one or two stochastic drivers, brief attention was paid to higher dimensional models. For example, we showed that the models developed in Proposition 1 can be extended to multi factor models with two stochastic drivers and up to 4 state variables. While algorithms for such models could be constructed, such an analysis might be premature. Indeed, since the class of two state variable models that now can be tested is large, significant empirical work needs to be conducted, and their limitations need to be identified, before their multifactor extensions are considered.

¹²This methodology is very well explained on Page 456 of Campbell et al (1996).

References

- Ait-Sahalia, Y., 1996, "Testing Continuous Time Models of the Spot Interest Rate," *Review of Financial Studies*, 9, 385-426
- Backus D., S. Foresi, and C. Telmer, 1997, "Models of Bond Pricing", unpublished paper, New York University.
- Billingsley, P. 1995, "Probability and Measure," John Willey and Sons, New York.
- Bollerslev, T, R. Chou and K. Kroner, 1992, "ARCH Modeling in Finance: A Review of the Theory and Empirical Evidence," *Journal of Econometrics* 52,1-59.
- Boyle, P., M. Broadie, and P. Glasserman, 1997, "Monte Carlo Methods for Security pricing," *Journal of Economic Dynamics and Control*, 21, 1267-1322.
- Brenner, R., R. Harjes, and K. Kroner, 1996, "Another Look at Alternative Models of the Short Term Interest Rate," *Journal of Financial and Quantitative Analysis*, 31, 85-107.
- Broadie, M. and P. Glasserman, 1997, "A Stochastic Mesh Method for Pricing High Dimensional American Options," working paper, Columbia University.
- Brown, R. and S. Schaefer, 1994, "The Term Structure of Real Interest Rates and the Cox, Ingersoll and Ross Model," *Journal of Financial Economics*," 35, 3-42.
- Brown, S. and P. Dybvig, 1986, "The Empirical Implications of the Cox, Ingersoll, Ross Theory of the Term Structure of Interest Rates," *Journal of Finance*, 41, 617-632.
- Campbell, J., A. Lo, C. MacKinlay, 1997, "The Econometrics of Financial Markets," Princeton Publishers.
- Chen, L., 1996, "Interest Rate Dynamics, Derivative pricing, and Risk Management", *Lecture Notes in Economics and Mathematical Systems*, 435, Springer.
- Chan, K., G. Karolyi, F. Longstaff, and A. Sanders, 1992, "An Empirical Comparison of Alternative Models of the Short Term Interest Rate," *Journal of Finance*, 47, 1209-1227.
- Constantinides, G., 1985, "A Theory of the Nominal Term Structure of Interest Rates," *The Review of Financial Studies*, 5, 531-552.
- Cox, J., J. Ingersoll, and S. Ross, 1985, "A Theory of the Term Structure of Interest Rates," *Econometrica*, 53, 363-384.
- Duan, J., 1997, "Augmented GARCH(p,q) Process and its Diffusion Limit," *Journal of Econometrics*, 79, 97-127.
- Duan, J., 1996b, "Term Structure and Bond Option Pricing under GARCH," , working paper, Hong Kong University of Science and Technology.

- Duffie, D., and R. Kan, 1996, "A Yield Factor Model of Interest Rates", *Mathematical Finance*, Vol. 6, 379-406.
- Gray, S., 1996, "Modeling the Conditional Distribution of Interest Rates as Regime Switching Processes," *Journal of Financial Economics*, 42, 27-62.
- Hamilton, J., 1989, "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57, 357-384.
- Heath, D., R. Jarrow, and A. Morton, 1992, "Bond pricing and The Term Structure of Interest Rates," *Econometrica*, 60, 77-105.
- Heston, S., and S. Nandi, 1999, "A Closed Form GARCH Option Pricing Model, working paper, Federal Reserve Bank of Atlanta.
- Inui, K. and M. Kijima, 1998, "A markovian Framework in Multi-Factor Heath-Jarrow-Morton Models," *Journal of Financial and Quantitative Analysis*, 33, 423-440.
- 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, "US Term Structure Data, 1947-1991," manuscript and computer diskettes, Ohio State University.
- Ritchken, P. and L. Sankarasubramanian, 1995, "Volatility Structures of Forward Rates and The Dynamics of The Term Structure," *Mathematical Finance*, 5, 55-72.
- Ritchken, P. and R. Trevor, 1999, "Pricing Options under Generalized GARCH and Stochastic Volatility Processes," *Journal of Finance*, 54, 377- 402
- Vasicek, O., 1977, "An Equilibrium Characterization of the Term Structure," *Journal of Financial Economics*," 5, 177-188.

Appendix 1

Proof of Lemma 1

We have

$$\frac{dQ}{dP} = \frac{m_T}{m_0} e^{\sum_{i=0}^{T-1} r_i}$$

Now, for any $t < T$, we have

$$E^P\left[\frac{dQ}{dP} \middle| \mathcal{F}_t\right] = \frac{m_t}{m_0} e^{\sum_{i=0}^{t-1} r_i}.$$

Under measure Q , if Z_{t+1} is any integrable \mathcal{F}_{t+1} measurable random variable, then:

$$E^Q[Z_{t+1} | \mathcal{F}_t] = E^P\left[Z_{t+1} e^{r_t} \frac{m_{t+1}}{m_t} \middle| \mathcal{F}_t\right]$$

To find the conditional distribution of ϵ_{t+1} under measure Q , note that:

$$E^Q[e^{c\epsilon_{t+1}} | \mathcal{F}_t] = E^P\left[e^{r_t} e^{c\epsilon_{t+1}} \frac{m_{t+1}}{m_t} \middle| \mathcal{F}_t\right]$$

Substituting for the marginal rate of substitution, and simplifying leads to:

$$E^Q[e^{c\epsilon_{t+1}} | \mathcal{F}_t] = e^{-\lambda_t c + \frac{c^2}{2}}$$

This implies that, under measure Q , $\epsilon_{t+1} \sim N(-\lambda_t, 1)$.

Similar computations show that under measure Q , $\eta_{t+1} \sim N(-\phi_t, 1)$

Proof of Proposition 1

The dynamics for the state variable under the Q measure is given by

$$\begin{aligned} r_{t+1} &= \alpha + \beta r_t + h_t(\epsilon_{t+1} - \lambda h_t) + \sigma(\epsilon_{t+1} - \lambda h_t)^2 \\ h_{t+1}^2 &= a_0 + a_1 h_t^2 + a_2(\epsilon_{t+1} - (\lambda + c)h_t)^2 + a_3(\eta_{t+1} - (\phi + e)h_t)^2 \end{aligned}$$

The proof is by induction. For $n = 1$

$$P(t, t+1) = e^{-r_t}$$

For $n = 2$,

$$\begin{aligned} P(t, t+2) &= e^{-r_t} E^Q[e^{-r_{t+1}} | \mathcal{F}_t] \\ &= e^{-r_t} E^Q\left[e^{-\alpha - \beta r_t + \lambda h_t^2 - h_t \epsilon_{t+1} - \sigma \epsilon_{t+1}^2 + 2\lambda \sigma h_t \epsilon_{t+1} - \sigma \lambda^2 h_t^2} \middle| \mathcal{F}_t\right] \end{aligned}$$

Computing this expectation leads to:

$$P(t, t+2) = e^{-A_2 - B_2 r_t + C_2 h_t^2}$$

Now, assume that

$$P(t, t+n) = e^{-A_n - B_n r_t + C_n h_t^2}.$$

Then,

$$P(t, t + n + 1) = e^{-r_t} E^Q[e^{-A_n - B_n r_{t+1} + C_n h_{t+1}^2} | \mathcal{F}_t]$$

which after simplifying works out to be:

$$P(t, t + n + 1) = e^{-A_{n+1} - B_{n+1} r_t + C_{n+1} h_t^2}.$$

Appendix 2: Two factor Models with 4 State Variables.

The results of Proposition 1 generalize to models with more than one stochastic driver. As an example, let the dynamics of the pricing kernel be:

$$-\ln\left(\frac{m_{t+1}}{m_t}\right) = \frac{\lambda_{1,t}^2}{2} + \frac{\lambda_{2,t}^2}{2} + x_{1,t} + x_{2,t} + \lambda_{1,t}\epsilon_{1,t+1} + \lambda_{2,t}\epsilon_{2,t+1}$$

where the state variables follow the processes:

$$\begin{aligned} x_{1,t+1} &= \alpha_1 + \beta_1 x_{1,t} + h_{1,t}\epsilon_{1,t+1} \\ x_{2,t+1} &= \alpha_2 + \beta_2 x_{2,t} + h_{2,t}\epsilon_{2,t+1} \end{aligned}$$

with

$$\begin{aligned} h_{1,t+1}^2 &= a_{10} + a_{11}h_{1,t}^2 + a_{12}(\epsilon_{1,t+1} - d_1 h_{1,t})^2 \\ h_{2,t+1}^2 &= a_{20} + a_{21}h_{2,t}^2 + a_{22}(\epsilon_{2,t+1} - d_2 h_{2,t})^2 \end{aligned}$$

Assuming the market prices of risk are proportional to their volatility levels we then can follow along identical lines as in Proposition 1, to obtain:

$$r_t = x_{1,t} + x_{2,t}$$

where valuation proceeds under the risk neutral measure:

$$\begin{aligned} x_{1,t+1} &= \alpha_1 + \beta_1 x_{1,t} - \lambda_{1,t} h_{1,t}^2 + h_{1,t}\epsilon_{1,t+1} \\ x_{2,t+1} &= \alpha_2 + \beta_2 x_{2,t} - \lambda_{2,t} h_{2,t}^2 + h_{2,t}\epsilon_{2,t+1} \end{aligned}$$

where

$$\begin{aligned} h_{1,t+1}^2 &= a_{10} + a_{11}h_{1,t}^2 + a_{12}(\epsilon_{1,t+1} - (d_1 + \lambda_1)h_{1,t})^2 \\ h_{2,t+1}^2 &= a_{20} + a_{21}h_{2,t}^2 + a_{22}(\epsilon_{2,t+1} - (d_2 + \lambda_2)h_{2,t})^2 \end{aligned}$$

The bond pricing equation then obtains as a product of two terms, with each term having a structure as in Proposition 1.

Appendix 3

Proof of Proposition 2

The proof is by induction. For one period, a simple application of the central limit theorem tells us that r_{t+1}^a converges in distribution to a normal random variable with the required mean and variance. Since h_{t+1}^2 is a function of the r_{t+1} and r_t , and h_{t+1}^2 has at most a finite number of discontinuities, an application of the continuous mapping theorem (see theorem 25.7 in Billingsley (1995)) shows that h_{t+1}^{2a} converges in distribution to h_{t+1}^2 . We now have to prove that the unconditional distributions across multiple periods converge. Now, if we show that the two period unconditional distributions converge, then the proof for multiple periods follows by induction. But the fact that two period unconditional distributions converge is a consequence of the following version of the continuous mapping theorem.

Lemma

Let X_n and Y_n be two sequences of random variables such that X_n converges in distribution to a random variable X and Y_n converges in distribution to Y . Suppose that $h : \mathbb{R}^2 \rightarrow \mathbb{R}^1$ is measurable and that the set D_h of its discontinuities is measurable. If $P[(X, Y) \in D_h] = 0$ then $h(X_n, Y_n)$ converges in distribution to $h(X, Y)$.

Proof of Proposition 3

Notice that we restrict the values of h_t to a finite number (K) of values. Using these values we approximate h_{t+1}^2 using simple step functions. Now, as we refine the partition (that is increase K) the approximating function closes the gap with the true function for h_{t+1}^2 . Using this observation, for any one period, the approximating function h_{next}^2 converges pointwise (that is for any given h_t) to the true function h_{t+1}^2 .

To prove that it converges across multiple periods, we first consider the convergence over two periods and then use induction to prove convergence over multiple periods. The two period short rate and volatility values are given by

$$\begin{aligned} r_{t+2} &= \alpha + \beta r_{t+1} + h_{t+1} r_{t+1}^\gamma (\epsilon_{t+2} - \lambda_{t+1}) \\ h_{t+2}^2 &= a_0 + a_1 h_{t+1}^2 + a_2 h_{t+1}^2 (\epsilon_{t+2} - (d + \lambda_t))^2 \end{aligned}$$

We see from these equations that both r_{t+2} and h_{t+2}^2 are continuous functions of h_{t+1}^2 which in turn are continuous functions of h_t^2 . Hence an approximation of h_{t+1}^2 by simple step functions which converges pointwise will also ensure that r_{t+2} and h_{t+2} converge pointwise.

Table 1: Convergence of Yields for the Vasicek Model.

This table reports the convergence of yields using the algorithm when the dynamics under the Q measure are given by

$$r_{t+1} = \alpha + \beta r_t + h_0(\epsilon_{t+1} - \lambda)$$

The model was implemented using a trinomial approximation for ϵ_{t+1} . The parameters were obtained by matching the moments as was done in Backus et.al (1997). In particular, they matched the true unconditional mean, variance and the auto-correlation of the one month short rate with the corresponding sample estimates obtained from the McCulloch-Kwon (1993) data that covers the period from January 1952 to February 1991, involving 470 monthly observations. This procedure yielded the values of α, β and h_0 . Finally, the value of λ was obtained by matching the theoretical yield with the average yield from the sample of the 10 year maturity bonds. The parameters obtained are $\alpha = 0.000106272$, $\beta = -0.024$, $\lambda = -0.082$, and $h_0 = 0.00055$ (or equivalently 0.67% per year). The initial short rate was taken to match the average short rate over the sample period, yielding $r_0 = 5.314\%$. Note that the scaling factor, h_0 , in this model remains unchanged over time. The table shows that the trinomial approximation with $m = 1$ is sufficient.

| | Maturity (Months) | | | | | | | | |
|------------|-------------------|--------|--------|--------|--------|--------|--------|--------|------|
| | 3 | 6 | 12 | 24 | 36 | 38 | 60 | 120 | Time |
| m=1 | 5.3682 | 5.4455 | 5.5869 | 5.8248 | 6.0157 | 6.1708 | 6.2981 | 6.6852 | 5.2 |
| Analytical | 5.3682 | 5.4455 | 5.5869 | 5.8249 | 6.0159 | 6.1710 | 6.2983 | 6.6856 | |

Table 2: Convergence of Yields for the GARCH-Vasicek Model.

This table reports the convergence of yields using the algorithm when the dynamics under the Q measure are given by

$$\begin{aligned} r_{t+1} &= \alpha + \beta r_t + h_t(\epsilon_{t+1} - \lambda) \\ h_{t+1}^2 &= a_0 + a_1 h_t^2 + a_2(\epsilon_{t+1} - \lambda h_t)^2 \end{aligned}$$

The values for α, β, r_0, h_0 are chosen as given in Table 1 and the rest of the parameters are $a_0 = 0.0000001$, $a_1 = 0.3$, $a_2 = 0.0000002$ and $\lambda h_0 = -0.082$. These values ensure that the long run average variance is stationary. In this example $m = 1$ and the table shows the convergence of yields as the number of regimes are increased to 30.

| | Maturity (Months) | | | | | | | | |
|------------|-------------------|-------|-------|-------|-------|-------|-------|-------|------|
| | 3 | 6 | 12 | 24 | 36 | 38 | 60 | 120 | Time |
| K=1 | 5.368 | 5.446 | 5.587 | 5.825 | 6.016 | 6.171 | 6.298 | 6.686 | 5.4 |
| K=10 | 5.374 | 5.471 | 5.661 | 5.983 | 6.242 | 6.452 | 6.624 | 7.147 | 41 |
| K=15 | 5.374 | 5.476 | 5.670 | 6.000 | 6.265 | 6.481 | 6.657 | 7.193 | 65 |
| K=20 | 5.373 | 5.473 | 5.667 | 5.598 | 6.264 | 6.480 | 6.657 | 7.193 | 86 |
| K=30 | 5.373 | 5.474 | 5.670 | 6.005 | 6.274 | 6.492 | 6.671 | 7.215 | 144 |
| Analytical | 5.373 | 5.475 | 5.671 | 6.004 | 6.273 | 6.490 | 6.669 | 7.212 | |

Table 3: Convergence of Yields for the CIR Model.

This table reports the convergence of yields using the algorithm when the dynamics under the Q measure are given by

$$r_{t+1} = \alpha + \beta r_t + h_0 r_t^\gamma (\epsilon_{t+1} - \lambda r_t^\gamma)$$

The model was implemented using a trinomial approximation for ϵ_{t+1} . As in the Vasicek model (see Table 1) the parameters were obtained by matching the theoretical moments with the sample moments of the one month short rate. The parameters are $\alpha = 0.000106272$, $\beta = -0.024$, $\lambda = -1.07$, and $h_0 = 0.00835577$. The initial spot rate was $r_0 = 5.314\%$. Note that the scaling factor, h_0 , in this model remains unchanged over time. The table shows that the trinomial approximation with $m = 1$ is sufficient.

| | Maturity (Months) | | | | | | | | |
|------------|-------------------|--------|--------|--------|--------|--------|--------|--------|------|
| | 3 | 6 | 12 | 24 | 36 | 38 | 60 | 120 | Time |
| M=1 | 5.3610 | 5.4288 | 5.5555 | 5.7771 | 5.9632 | 6.1201 | 6.2532 | 6.6859 | 6 |
| Analytical | 5.3610 | 5.4288 | 5.5555 | 5.7772 | 5.9633 | 6.1203 | 6.2535 | 6.6862 | |

Table 4: Convergence of Yields for the GARCH-CIR Model.

This table reports the convergence of yields using the algorithm when the dynamics under the Q measure are given by

$$r_{t+1} = \alpha + \beta r_t + h_0 r_t^\gamma (\epsilon_{t+1} - \lambda r_t^\gamma)$$

$$h_{t+1}^2 = a_0 + a_1 h_t^2 + a_2 (\epsilon_{t+1} - \lambda r_t^\gamma)^2$$

The values for α , β , r_0 , h_0 are chosen as given in Table 3 and the rest of the parameters are $a_0 = 0.000005$, $a_1 = 0.3$, $a_2 = 0.68$ and $\lambda\sqrt{r_0} = -9.862$. In this example $m = 1$ and the table shows the convergence of yields as the number of regimes increase to 40.

| | Maturity (Months) | | | | | | | | |
|------|-------------------|-------|-------|-------|-------|-------|-------|-------|------|
| | 3 | 6 | 12 | 24 | 36 | 38 | 60 | 120 | Time |
| K=1 | 5.361 | 5.429 | 5.555 | 5.777 | 5.963 | 6.120 | 6.253 | 6.686 | 6.2 |
| K=10 | 5.358 | 5.417 | 5.526 | 5.718 | 5.879 | 6.016 | 6.132 | 6.511 | 36 |
| K=20 | 5.358 | 5.412 | 5.506 | 5.660 | 5.786 | 5.892 | 5.981 | 6.268 | 78 |
| K=25 | 5.358 | 5.411 | 5.502 | 5.648 | 5.768 | 5.868 | 5.953 | 6.227 | 97 |
| K=30 | 5.358 | 5.410 | 5.498 | 5.640 | 5.756 | 5.854 | 5.936 | 6.204 | 139 |
| K=40 | 5.358 | 5.410 | 5.497 | 5.636 | 5.754 | 5.848 | 5.925 | 6.194 | 193 |

Table 5: Convergence of Yields for Model with Chi-Squared Innovations.

This table reports the convergence of yields using the algorithm when the dynamics under the Q measure are given by

$$r_{t+1} = \alpha + \beta r_t + h_0(\epsilon_{t+1} - \lambda) + \sigma(\epsilon_{t+1} - \lambda)^2$$

The parameters used for this model are $\alpha = 0.00010628$, $\beta = 0.8$, $\lambda = -0.0824$, $r_0 = 5.314\%$, $h_0 = 0.000278$ and $\sigma = 0.001$. The gap, Δ , between successive interest rate is 20 basis points. The table shows the convergence of prices to the true theoretical values.

| | Maturity (Months) | | | | | | | | |
|------------|-------------------|-------|-------|-------|-------|-------|-------|-------|------|
| | 3 | 6 | 12 | 24 | 36 | 48 | 60 | 120 | Time |
| m=1 | 5.589 | 5.874 | 6.189 | 6.433 | 6.525 | 6.572 | 6.600 | 6.655 | 13 |
| m=2 | 5.591 | 5.879 | 6.201 | 6.453 | 6.548 | 6.596 | 6.625 | 6.683 | 17 |
| m=3 | 5.592 | 5.883 | 6.212 | 6.471 | 6.569 | 6.619 | 6.649 | 6.708 | 22 |
| m=4 | 5.592 | 5.885 | 6.216 | 6.478 | 6.577 | 6.627 | 6.657 | 6.717 | 26 |
| Analytical | 5.592 | 5.886 | 6.218 | 6.482 | 6.581 | 6.632 | 6.662 | 6.723 | |

Figure 1: Distribution of r_t after 12 time periods for the GARCH-Vasicek model
 The densities were obtained by simulating the following model for different parameter values.

$$\begin{aligned} r_{t+1} &= \alpha + \beta r_t + h_t \epsilon_{t+1} \\ h_{t+1}^2 &= a_0 + a_1 h_t^2 + a_2 (\epsilon_{t+1} - dh_t)^2 \end{aligned}$$

The benchmark parameters in case **1** are: $\alpha = 0.000106272$, $\beta = 0.976$, $a_0 = 0$, $a_1 = 0.3$, $d = 0$ and $a_2 = 3.09E-09$. The starting values for r_0 and h_0 were set at $r_0 = 6.3\%$, $h_0 = 0.00055$ (or equivalently 0.67% per year). For cases **2** and **3** the value of d was set at 15000 and -15000 respectively. As can be seen for different parameters a wide range of skewness can be observed by controlling the level of just one parameter.

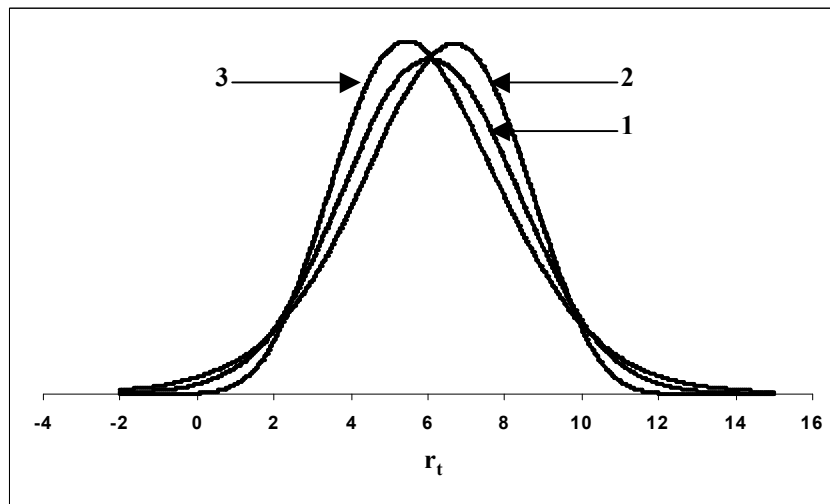


Figure 2: Convergence of Option Prices in the GARCH-Vasicek Model

This figure shows the convergence of option prices in the GARCH-CIR model. The model used for computing the option prices and the corresponding parameter values are the same as given in table 2. The option considered here is a 1 year call option on a 2 year bond. The strike price was set at the forward price.

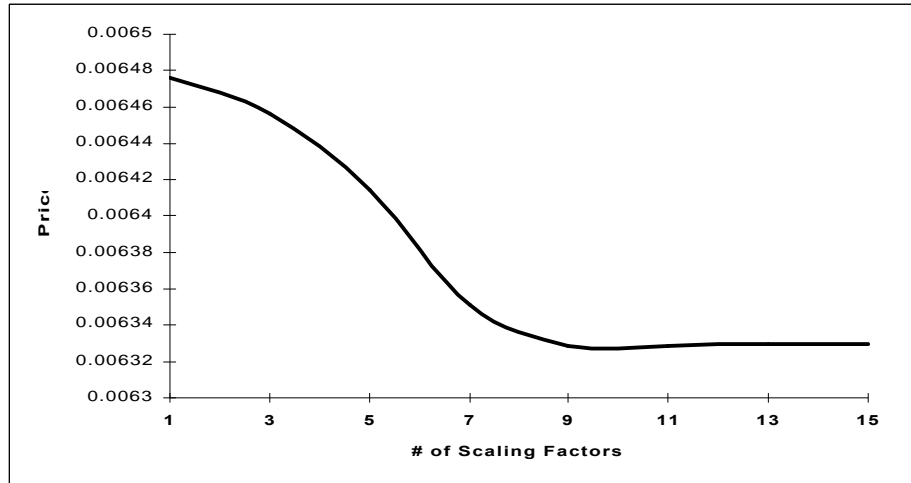


Figure 3: Convergence of Option Prices in the GARCH-CIR Model

This figure shows the convergence of option prices in the GARCH-CIR model. The model used for computing the option prices and the corresponding parameter values are the same as given in table 4. The option considered here is a 1 year call option on a 2 year bond. The strike price was set at the forward price.

