Modeling the conditional distribution of interest rates as a regime-switching process

Stephen F. Gray

Fuqua School of Business, Duke University, Durham NC 27708, USA

Abstract

This paper develops a generalized regime-switching (GRS) model of the short-term interest rate. The model allows the short rate to exhibit both mean reversion and conditional heteroskedasticity and nests the popular generalized autoregressive conditional heteroskedasticity (GARCH) and square root process specifications. The conditional variance process accommodates volatility clustering *and* dependence on the level of the interest rate. A first-order Markov process with *state-dependent* transition probabilities governs the switching between regimes. The GRS model is compared with various existing models of the short rate in terms of (1) the statistical fit of short-term interest rate data and (2) out-of-sample forecasting performance.

Key words: Short-term interest rates; Regime-switching; Conditional volatility; Maximum likelihood estimation
JEL classification: G12

1. Introduction

The key role that the nominal short-term interest rate plays in the valuation of almost all securities has made it one of the most frequently modeled variables in financial economics. Unfortunately, many popular models of the short

rate produce untenable results when fit to the data. For example, stationarity is only guaranteed for special types of diffusion models, and estimates of generalized autoregressive conditional heteroskedasticity (GARCH) models often imply an explosive conditional variance. These issues have implications for forecasting volatility which is important for valuing interest-rate-sensitive securities and interest rate risk management. While existing models of the short rate have many appealing features, parameter estimates that imply explosive variances, for example, can present serious problems in certain applications. This paper explores the idea that, while the basic structure of existing models provides a good characterization of the short rate process, time variation in the parameter values may produce untenable results. This is done by developing a generalized regime-switching (GRS) model that nests many existing interest rate models as special cases and allows parameter values to change over time with changes in regime.

One potential source of misspecification of existing models of the short rate is that the structural form of conditional means and variances is relatively inflexible and is held fixed throughout the entire sample period. These models are *single-regime* models in the sense that they effectively assume a single structure for the conditional mean and variance. In this paper, the assumption of a single regime is relaxed in favor of a *regime-switching* model of the short-term interest rate. The coefficients in this model are different in each regime to account for the possibility that the economic mechanism that generates the short rate may undergo a finite number of changes over the sample period. These coefficients are unknown and must be estimated, and, although the regimes are never observed, probabilistic statements can be made about the relative likelihood of their occurrence, conditional on an information set.

Ultimately, of course, all such models involve the estimation of a set of parameters that are assumed to be fixed over the sample period. In this sense, a regime-switching model is a more complex, flexible structure, able to model data generated by different economic mechanisms – all within a single, unified model. For example, a single-regime model may assume that the short rate is mean-reverting. This amounts to assuming that the long-run mean and the speed of reversion are the same throughout the sample. In contrast, a regime-switching model is flexible enough to incorporate a different speed of reversion to a different long-run mean at different times throughout the sample period. Introducing the possibility of regime switches is one way of relaxing the linearity (of the long-run mean and speed of reversion) inherent in most single-regime models. Regime-switching models thus have the attractive feature of incorporating significant nonlinearities, while remaining tractable and easy to estimate.

The Federal Reserve (Fed) experiment of 1979 to 1982 has, in part, motivated previous work on regime-switching models of interest rates (see Hamilton, 1988; Cai, 1994). During this period, the Fed deviated from its usual practice of targeting interest rates and experimented with using nonborrowed reserves (NBR's) as a new target instrument for monetary policy. The result was a period of

unprecedented interest rate volatility. This change in operating procedure is likely to have affected the structure of the dependence of nominal interest rates on the set of explanatory variables. Indeed, it is probably unreasonable to expect that the structural relation between the various variables would be preserved. Other periods of high volatility in U.S. interest rates have also coincided with changes in the economic environment due to the OPEC oil crisis, the October 1987 stock market crash, and wars involving the U.S.

The generalized regime-switching model developed in this paper advances the regime-switching literature in a number of directions. In most existing regime-switching models, the moments, conditional on the regime, are constant. The mean and variance of the short rate are usually held constant within each regime. This rules out within-regime mean reversion or conditional heteroskedasticity. The model developed here accommodates time-varying conditional moments – mean reversion and conditional heteroskedasticity within each regime. Moreover, the model goes beyond the Markov autoregressive conditional heteroskedasticity (ARCH) models of Cai (1994) and Hamilton and Susmel (1994) in that the conditional variances are made flexible enough to incorporate the important persistence associated with GARCH effects. In contrast to this earlier work, all of the GARCH parameters (rather than just an additive or multiplicative scaling parameter) are regime-dependent. This is an important feature if, as is likely to be the case, the persistence of individual shocks is lower during periods of extreme volatility. Friedman and Laibson (1989), for example, report that large shocks to stock market returns are not persistent, but that more moderate shocks are quite persistent. Also, one can imagine models in which a large shock has the effect of 'relieving pressure' on the system, in which case very large shocks are less persistent than more moderate shocks. For example, consider a realignment in the European Monetary System – a very large shock to the exchange rate which is typically followed by lower volatility. Bekaert and Gray (1995) incorporate this feature in a model of exchange rates.

The conditional variance in each regime can also depend upon the level of interest rates. Thus, the conditional variance process accommodates volatility clustering *and* dependence on the level of the interest rate, nesting each individual specification as a special case. Moreover, the Markov transition probabilities are state-dependent, being a function of the level of interest rates. This allows for the possibility that a switch to the high-volatility regime may be more likely when interest rates are high, for example. In most existing regime-switching models, the regime-switching probabilities (i.e., the probability of switching from one particular regime to another) are constant although preliminary work on time-varying probabilities has been done in a variety of settings (see, for example, Diebold, Lee, and Weinbach, 1994; Ghysels, 1993; Filardo, 1993, 1994). Finally, the paper makes use of the recursive nature of the regime probabilities developed in Gray (1995) and Hamilton (1994) in simplifying construction of the likelihood function. This permits easy estimation of relatively complicated models.

The GRS model is shown to deliver sensible results, capturing the features of short-term interest rate data better than existing models in terms of (1) the statistical fit of short-term interest rate data and (2) the of out-of-sample forecasting performance.

The paper is organized as follows. The next section reviews existing interest rate models and some of the stylized facts about short-term interest rates. Section 3 presents the generalized regime-switching model. Section 4 discusses estimation issues and contains an empirical analysis of the GRS model as well as several popular models nested within the GRS framework. Section 5 highlights the economic implications of the results. Section 6 summarizes.

2. Existing interest rate models

Models of the short-term interest rate must capture two well-known empirical attributes. First, the short rate is mean-reverting. The simplest and most common way of modeling mean reversion is to let next period's change in the short rate depend linearly on the current short rate level. Second, the unconditional distribution of changes in the short rate is leptokurtic. Engle (1982) shows that a possible cause of the leptokurtosis in the unconditional distribution is conditional heteroskedasticity, and two different ways of modeling this conditional heteroskedasticity are common in the literature. In various continuous time or diffusion models, the conditional variance of changes in the short rate is a function of the level of the short rate. In the ARCH/GARCH models of Engle (1982) and Bollerslev (1986), the conditional variance of changes in the short rate depends upon lagged squared shocks to the short rate. For many empirical applications, however, conditional heteroskedasticity modeled in these ways is insufficient to account for the degree of leptokurtosis in the data. The two most common classes of interest rate models are reviewed below.

2.1. Diffusion models

In continuous time or diffusion models, the short rate is usually based on Brownian motion. In a general framework, the dynamics of the short rate can be described by the stochastic differential equation:

$$dr = (\alpha + \beta r)dt + \sigma r^{\gamma}dW , \tag{1}$$

where dW is the increment from a standard Brownian motion. Chan, Karolyi, Longstaff, and Sanders (1992) note that this equation encompasses a number of models used in the literature. In particular, where $\gamma = 0.5$, the stochastic differential equation describes the square root process used by Cox, Ingersoll, and Ross (1985). For estimation purposes, the stochastic differential Eq. (1) is usually

discretized as in Chan et al.:

$$\Delta r_t = r_t - r_{t-1} = \alpha + \beta r_{t-1} + \varepsilon_t \,, \tag{2}$$

where $E[\varepsilon_t|\Phi_{t-1}] = 0$ and $E[\varepsilon_t^2|\Phi_{t-1}] = \sigma^2 r_{t-1}^{2\gamma}$, and Φ_{t-1} is the agents' information set at time $t-1$. Mean reversion can be captured by setting $\beta < 0$ and leptokurtosis by introducing conditional heteroskedasticity (setting $\gamma > 0$). This discretization serves to approximate the true process. In some special cases, for example the constant elasticity of variance (CEV) process, the square root process, and the constant volatility process, the continuous time likelihood can be estimated directly so that discretization is unnecessary.

2.2. GARCH models

Bollerslev, Chou, and Kroner (1992) review a number of interest rate applications of ARCH/GARCH models. The GARCH (1,1) model takes the following form:

$$\Delta r_t = X_{t-1}\beta + \varepsilon_t \,, \tag{3}$$

$$\varepsilon_t|\Phi_{t-1} \sim N(0, h_t)\,,$$

$$h_t = \omega + a\varepsilon_{t-1}^2 + bh_{t-1} \,.$$

Note that X_{t-1} in (3) may include r_{t-1} to capture mean reversion. The introduction of conditional heteroskedasticity can capture leptokurtosis in the unconditional distribution.

One problem with fitting GARCH models to short-term interest rates is that the parameter estimates often imply that the conditional variance process is not covariance-stationary. Bollerslev (1986) demonstrates that this is the case when $a + b$ is greater than one. Nelson (1990) shows that in such cases shocks accumulate in the sense that $E[h_{t+m}|h_t] \rightarrow \infty$ as $m \rightarrow \infty$. Note, however, that this does not imply that $h_{t+m} \rightarrow \infty$ as $m \rightarrow \infty$.

Engle, Ng, and Rothschild (1990) report $a_1 + b_1 = 1.0096$ for a portfolio of U.S. Treasury Securities. Kees, Nissen, Schotman, and Wolff (1994) report $a + b = 1.10$ for one-month T-bills, Hong (1988) reports $a + b = 1.073$ for excess returns on three-month T-bills over one-month T-bills, and Engle, Lilien, and Robbins (1987) report similar results using an ARCH (12) model on quarterly data for the excess holding yield of six-month T-bills over three-month T-bills. Frequently, however, statistical tests are unable to reject the hypothesis that the conditional variance of the short rate follows an integrated GARCH (IGARCH) process ($a + b = 1$). In any event, we argue below that this strong persistence in volatility may be an artifact of changes in the economic mechanism generating the short rate. Lamoureux and Lastrapes (1990), for example, demonstrate that any structural shift in the unconditional variance is likely to lead to misestimation of the GARCH parameters such that they imply too much persistence in volatility.

That is, regime shifts are mistaken for periods of volatility clustering. While Lamoureux and Lastrapes restrict their analysis to shifts in the unconditional variance, all GARCH parameters may vary across regimes. The ARCH parameter (a) may be driven up by occasional periods of dramatic volatility with low persistence, for example, while the GARCH parameter (b) remains high due to the strong persistence in less dramatic periods. These issues are now examined formally in a GRS framework.

3. Model development

This section outlines the generalized regime-switching model. The GRS model examined in this paper nests the GARCH (1,1) model, a discretized diffusion model motivated by the Cox, Ingersoll, and Ross (1985) model, and the Markov regime-switching models that have appeared in the literature to date. Although not explored in this paper, the GRS framework can be extended to accommodate other conditional variance processes. The exponential GARCH (EGARCH) model of Nelson (1991), for example, can be accommodated by modeling the *log* of the conditional variance as being different across regimes. The actual conditional variance is recovered by taking the exponent. Moreover, the conditional variance in each regime could be parameterized as taking a flexible form such as that examined by Hentschel (1995).

Both the discretized diffusion and GARCH models take the following form:

$$\Delta r_t = \mu(\theta_\mu, \Phi_{t-1}) + \sqrt{h(\theta_h, \Phi_{t-1})} z_t$$
$$= \mu_t + \sqrt{h_t} z_t , \tag{4}$$

where θ_μ and θ_h are vectors of unknown parameters and the z_t's are independent and identically distributed with mean zero and unit variance. In both models

$$\mu(\theta_\mu, \Phi_{t-1}) = \mu_t = \alpha + \beta r_{t-1} ,$$

where $\theta_\mu = \{\alpha, \beta\}$. In the discretized diffusion model, the conditional variance is

$$h(\theta_h, \Phi_{t-1}) = h_t = \sigma^2 r_{t-1}^{2\gamma} ,$$

where $\theta_h = \{\sigma^2, \gamma\}$, while in the GARCH model, the conditional variance is

$$h(\theta_h, \Phi_{t-1}) = h_t = \omega + a\varepsilon_{t-1}^2 + bh_{t-1} ,$$

where $\theta_h = \{\omega, a, b\}$. In the case of the discretized diffusion model, the only relevant conditioning information is the level of interest rates, so $\Phi_{t-1} = \{r_{t-1}\}$. The recursive nature of the GARCH model, however, makes the conditional variance dependent on the entire past history of the data so that $\Phi_{t-1} = \tilde{r}_{t-1}$ where $\tilde{r}_{t-1} = \{r_{t-1}, r_{t-2}, \ldots\}$.

The GRS model is a *generalized* regime-switching model in the sense that it is more general than any existing regime-switching model of the short rate along several dimensions outlined below. Each regime has a different degree of mean reversion to a different long-run mean. The conditional variance in each regime takes a very general form incorporating GARCH (not just ARCH) effects *and* level effects consistent with a square root process. The switching probabilities are time-varying, depending on the level of the short rate. The model has the following general form:

$$\Delta r_t = \mu[\theta_\mu(S_t), \Phi_{t-1}] + \sqrt{h[\theta_h(S_t), \Phi_{t-1}]} z_t , \tag{5}$$

where S_t is the *unobserved* regime at time t. Note that Φ_{t-1} does not contain S_t or lagged values of S_t. In this paper S_t takes one of two values (1 or 2), although in principle the methodology can be extended to accommodate (finitely) more regimes. For notational convenience, rewriting the above expression yields

$$\Delta r_t = \mu_{it} + \sqrt{h_{it}} z_t ,$$

when $S_t = i$.

The basic idea underlying the GRS model is simple. The parameters of the conditional mean and conditional variance process are allowed to take two different values, which depend on the value of the latent regime indicator S_t. The model can easily be generalized to allow not only the parameters but also the functional forms to vary over regimes. If conditional normality is assumed for each regime, for example, the conditional distribution of Δr_t is a mixture of distributions, that is $N(\mu_{1t}, h_{1t})$ in regime 1 and $N(\mu_{2t}, h_{2t})$ in regime 2, which can be written as

$$\Delta r_t | \Phi_{t-1} \sim \begin{cases} N(\mu_{1t}, h_{1t}) & \text{w.p.} \quad p_{1t} , \\ N(\mu_{2t}, h_{2t}) & \text{w.p.} \quad (1 - p_{1t}) , \end{cases} \tag{6}$$

where $p_{1t} = \Pr(S_t = 1 | \Phi_{t-1})$.

3.1. Specification of the conditional means

In the most general version of the model examined in this paper, the functional form of the conditional mean incorporates mean reversion in the standard way so that

$$\mu_{it} = \alpha_i + \beta_i r_{t-1} . \tag{7}$$

Within this framework, the conditional mean and variance could have an even more general parameterization. For example, the means could be autoregressive moving average, ARMA(p,q), and the variances could be GARCH(p,q). The parameterization adopted here, represents a tradeoff between flexibility and parsimony.

3.2. Specification of the conditional variances

The empirical evidence on the volatility of the short-term interest rate suggests that two factors are important. First, large (small) changes tend to be followed by large (small) changes. This volatility clustering is usually captured by GARCH-type models. Second, volatility tends to be higher when the short rate is high. This level effect is usually captured by diffusion-type models.

To model the conditional variance function, a specification that incorporates both GARCH and level effects is needed. Both Cai (1994) and Hamilton and Susmel (1994) have argued that regime-switching GARCH models are essentially intractable and impossible to estimate due to the dependence of the conditional variance on the entire past history of the data in a GARCH model. That is, the distribution at time t, conditional on the regime (S_t) and on available information (Φ_{t-1}), depends directly on S_t, and also indirectly on $\{S_{t-1}, S_{t-2}, \ldots\}$ due to the path dependence inherent in regime-switching GARCH models. This is because the conditional variance at time t depends upon the conditional variance at time $t-1$, which depends upon the regime at time $t-1$ and on the conditional variance at time $t-2$, and so on. Consequently, the conditional variance at time t depends on the entire sequence of regimes up to time t. The likelihood function is constructed by integrating over all possible paths. For the tth observation in a K-regime model, there are K^t components of the likelihood function, rendering estimation intractable for large sample sizes. This path dependence is illustrated in Fig. 1. To avoid this problem, Cai (1994) and Hamilton and Susmel (1994) develop regime-switching models in which the conditional variance in each regime is characterized by a low-order ARCH process.

The problem of path dependence, however, can be solved in a way that preserves the essential nature of the GARCH process (including the important persistence effects) yet allows tractable estimation of the model. Recall from Eq. (6) that in the GRS model, the conditional density of the short rate is essentially a mixture of distributions with time-varying mixing parameter. If conditional normality is assumed within each regime, the variance of changes in the short rate at time t is given by

$$
\begin{aligned}
h_t &= E[\Delta r_t^2 | \Phi_{t-1}] - E[\Delta r_t | \Phi_{t-1}]^2 \\
&= p_{1t}(\mu_{1t}^2 + h_{1t}) + (1 - p_{1t})(\mu_{2t}^2 + h_{2t}) - [p_{1t}\mu_{1t} + (1 - p_{1t})\mu_{2t}]^2 . \quad (8)
\end{aligned}
$$

Now h_t, which is not path-dependent, can be used as the lagged conditional variance in constructing h_{1t+1} and h_{2t+1} which follow GARCH processes. That is,

$$
h_{it} = \omega_i + a_i \varepsilon_{t-1}^2 + b_i h_{t-1} ,
$$

$$
h_{t-1} = p_{1t-1}[\mu_{1t-1}^2 + h_{1t-1}] + (1 - p_{1t-1})[\mu_{2t-1}^2 + h_{2t-1}]
$$

$$
- [p_{1t-1}\mu_{1t-1} + (1 - p_{1t-1})\mu_{2t-1}]^2 ,
$$

$$h_{2|1,1} = \omega_1 + a_1 \varepsilon_{1|1} + b_1 h_{1|1}$$

$$h_{1|1} = \omega_1 + a_1 \varepsilon_0^2 + b_1 h_0$$

$$h_{2|1,2} = \omega_2 + a_2 \varepsilon_{1|1} + b_2 h_{1|1}$$

$$h_0$$

$$h_{2|2,1} = \omega_1 + a_1 \varepsilon_{1|2} + b_1 h_{1|2}$$

$$h_{1|2} = \omega_2 + a_2 \varepsilon_0^2 + b_2 h_0$$

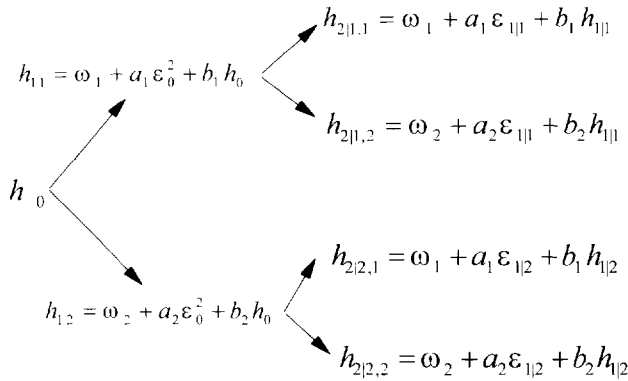$$h_{2|2,2} = \omega_2 + a_2 \varepsilon_{1|2} + b_2 h_{1|2}$$

Fig. 1. This figure illustrates the evolution of conditional variances in a path-dependent GARCH model. Each conditional variance depends not just on the current regime but on the entire past history of the process since the tree is not recombining. This is the literal extension of the Markov-ARCH models in Cai (1994) and Hamilton and Susmel (1994) to incorporate the persistence associated with GARCH effects.

The evolution of regimes is made precise in the subscripts to the conditional variances so that $h_{2|1,2}$, for example, represents the conditional variance at time 2, given that the process was in regimes 1 and 2, respectively, at times 1 and 2. Similarly, $\varepsilon_{1|2}$, for example, represents the unexpected change in the short rate at time 1, given that the process was then in regime 2. ω_i, a_i, and b_i, $i = 1, 2$, are unknown parameters to be estimated.

$$\varepsilon_{t-1} = \Delta r_{t-1} - E[\Delta r_{t-1} | \Phi_{t-2}]$$

$$= \Delta r_{t-1} - [p_{1t-1} \mu_{1t-1} + (1 - p_{1t-1}) \mu_{2t-1}].$$

Figs. 1 and 2 show the difference between the path-dependent approach of Cai (1994) and Hamilton and Susmel (1994) and the non-path-dependent approach developed above. In the path-dependent approach, h_{1t} (the conditional variance when $S_t = 1$) is different if the process is staying in regime 1 ($S_{t-1} = 1$) than if the process is switching from regime 2 ($S_{t-1} = 2$). Further, the conditional variances at time $t - 1$ depend on which regime the process was in at time $t - 2$ and so on. In the regime-switching GARCH model developed above, this path dependence is removed by aggregating the conditional variances from the two regimes at each time step. This single aggregated conditional variance (conditional on available information, but aggregated over regimes) is then all that is required to compute the conditional variances at the next time step. In addition to incorporating GARCH effects, the conditional variance specification in the GRS model also incorporates level effects as in the square-root process of Cox, Ingersoll, and Ross (1985). In particular, the most general specification is

$$h_{it} = \omega_i + a_i \varepsilon_{t-1}^2 + b_i h_{t-1} + \sigma_i^2 \sqrt{r_{t-1}}, \qquad (9)$$

$$h_{11} = \omega_1 + a_1 \varepsilon_0^2 + b_1 h_0$$

$$h_{2|1} = \omega_1 + a_1 \varepsilon_1^2 + b_1 h_1$$

$$h_1 = p_1(\mu_{1|1}^2 + h_{11}) + (1 - p_1)(\mu_{1|2}^2 + h_{1|2})$$

$$- \left[p_1 \mu_{1|1} + (1 - p_1)\mu_{1|2} \right]^2$$

$$\varepsilon_1 = y_1 - \left[p_1 \mu_{1,1} + (1 - p_1)\mu_{1\,2} \right]$$

$$h_0$$

$$h_{1|2} = \omega_2 + a_2 \varepsilon_0^2 + b_2 h_0$$

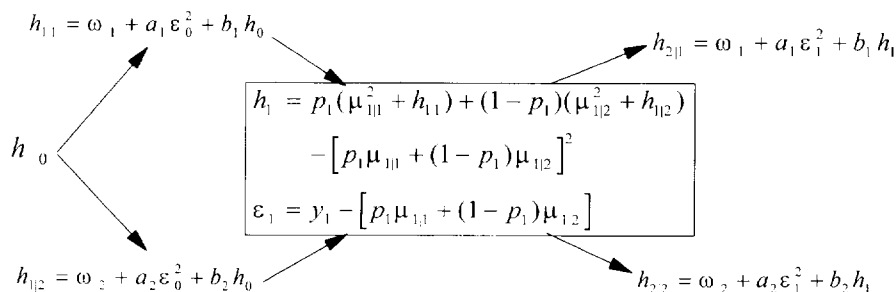$$h_{2\,2} = \omega_2 + a_2 \varepsilon_1^2 + b_2 h_1$$

Fig. 2. This figure illustrates the evolution of conditional variances in a path-independent GARCH model. Each conditional variance depends only on the current regime, not on the entire past history of the process. since the tree is recombining.

The conditional variances depend not on the evolution of regimes but only on the current regime so that $h_{2|1}$, for example, represents the conditional variance at time 2, given that the process is then in regime 1. Similarly, $\varepsilon_{1\,2}$, for example, represents the unexpected change in the short rate at time 1, given that the process was then in regime 2. At each point in time, dependence on the regime can be 'integrated out' by summing over all possible regimes to construct the variance conditional on observable information but not on the regime. For example, h_1 represents the variance of changes in the short rate at time 1 conditional on observable information. $\mu_{1|2}$, for example, is the expected change in interest rates at time 1 given that the process is then in regime 2 and p_1 is the probability that the process is in regime 1 at time 1, conditional on available information. ω_i, a_i, and b_i, $i = 1,2$, are unknown parameters to be estimated.

where ω_i, a_i, b_i, and σ_i, $i = 1,2$, are unknown parameters to be estimated.

Note that γ in Eq. (1) is set to 0.5 rather than estimated as a free parameter. This is because γ and σ are *highly* collinear, rendering interpretation of individual parameter estimates questionable at best.[1] Furthermore, fixing $\gamma = 0.5$ avoids potential nonstationarity problems associated with $\gamma > 1$ and facilitates interpretation in terms of the CIR model. Moreover, since γ is scale-invariant and σ is not, it makes more sense to fix γ and estimate σ rather than the reverse.

3.3. Specification of the switching probabilities

The only remaining structure to be placed on the model is the parameterization of the latent regime indicator S_t. Following Hamilton (1988, 1989, 1990), S_t can be parameterized as a first-order Markov process. The most common approach

---

[1] Estimating the unconstrained models produced correlations between γ and σ in the order of -0.98 to -0.99. Ball and Torous (1994), in addressing a similar issue, report that the likelihood function is very flat in $\gamma - \sigma$ space such that different $\gamma - \sigma$ pairs result in similar likelihood values.

in the literature is to use a constant matrix of transition probabilities:

$$
\begin{aligned}
\Pr[S_t = 1 | S_{t-1} = 1] &= P, \\
\Pr[S_t = 2 | S_{t-1} = 1] &= (1 - P), \\
\Pr[S_t = 2 | S_{t-1} = 2] &= Q, \\
\Pr[S_t = 1 | S_{t-1} = 2] &= (1 - Q).
\end{aligned}
\tag{10}
$$

The recursive nature of this Markov structure, however, can be exploited in such a way as to make the extension to time-varying transition probabilities straightforward. In the GRS model, the switching probabilities are dependent on the level of the short rate. In particular, $P_t = \Phi(c_1 + d_1 r_{t-1})$ and $Q_t = \Phi(c_2 + d_2 r_{t-1})$, where c_i and d_i, $i = 1, 2$, are unknown parameters and $\Phi(\cdot)$ is the cumulative normal distribution function which ensures that $0 < P_t, Q_t < 1$. Thus the GRS model takes the form of Eq. (6), with conditional means in Eq. (7), conditional variances in Eq. (9), and transition probabilities P_t and Q_t.

Finally, note that both mean reversion and leptokurtosis can be caused by the switches between regimes if the switching probabilities are correlated with r_{t-1}. To see this, suppose that changes in short-term interest rates are parameterized as being distributed $N(\alpha_i, \sigma_i^2)$ in regime i. Mean reversion exists if $\text{cov}(\Delta r_t, r_{t-1}) < 0$. But

$$
\begin{aligned}
\text{cov}(r_t - r_{t-1}, r_{t-1}) &= \text{cov}(E[r_t - r_{t-1} | r_{t-1}], r_{t-1}) \\
&= \text{cov}(p_{1t}\alpha_1 + (1 - p_{1t})\alpha_2, r_{t-1}) \\
&= (\alpha_1 - \alpha_2)\text{cov}(p_{1t}, r_{t-1}).
\end{aligned}
$$

Hence, if the regime probability (p_{1t}) is correlated with the level of interest rates, then switches between regimes may drive the observed mean reversion in the short rate. A similar argument applies with respect to conditional heteroskedasticity.

4. Estimation and results

The regime-switching literature typically uses (quasi) maximum likelihood (QML) estimation (the consistency and asymptotic normality of QML estimates, for some regime-switching models, under relatively mild regularity conditions is proven in Gray, 1995). The likelihood function for the GRS model is derived in the Appendix. For the most general model examined in this paper, μ_{it} is defined in (7), h_{it} is defined in (9), and

$$
\begin{aligned}
p_{1t} = (1 - Q_t) &\left[\frac{g_{2t-1}(1 - p_{1t-1})}{g_{1t-1}p_{1t-1} + g_{2t-1}(1 - p_{1t-1})} \right] \\
&+ P_t \left[\frac{g_{1t-1}p_{1t-1}}{g_{1t-1}p_{1t-1} + g_{2t-1}(1 - p_{1t-1})} \right],
\end{aligned}
\tag{11}
$$

where

$$p_{1t} = \Pr(S_t = 1 | \tilde{r}_{t-1}),$$

$$g_{1t} = f(\Delta r_t | S_t = 1),$$

$$g_{2t} = f(\Delta r_t | S_t = 2),$$

$$P_t = \Phi(c_1 + d_1 r_{t-1}),$$

$$Q_t = \Phi(c_2 + d_2 r_{t-1}).$$

The terms in square brackets in (11) represent $\Pr[S_{t-1} = 2 | \Phi_{t-1}]$ and $\Pr[S_{t-1} = 1 | \Phi_{t-1}]$, respectively.

The log-likelihood function can be constructed recursively, in the same way as in a GARCH model. This is done by choosing a 'startup' value or initial conditions for the probability process. This value can either be (a) estimated as an additional parameter or (b) set at some arbitrary value (say the unconditional probability) on the basis that the initial choice of the probability process becomes irrelevant in a large sample. The whole series of p_{1t}'s can then be built up recursively in the same way that the conditional variance process is built up recursively in GARCH estimation. The recursive structure of p_{1t} makes the regime-switching model similar to a GARCH model. In a GARCH model, successive innovations are drawn from distributions with different (conditional) variances, and these unobserved conditional variances follow a recursive structure with unknown (estimated) parameters. A similar interpretation – an unobserved parameter of the conditional distribution, following a recursive structure – can be given to p_{1t} here. The T-vector of likelihood function values can be filled in recursively in exactly the same way as for a GARCH model. This simplifies the conditioning process as all of the information in the past sequence of regimes is effectively summarized by p_{1t-1}.

4.1. Data

The data used in this study are one-month U.S. Treasury bill rates obtained from the Federal Reserve and CRB Infotech database. The weekly data set consists of annualized yields recorded at noon Eastern Standard Time every Wednesday for the period January 1970 through April 1994, a total of 1,267 observations. Fig. 3 plots the data, illustrating the dramatic changes in interest rates that occurred during the Fed experiment and the OPEC oil crises. Fig. 3, lower panel, contains plots of weekly changes in short-term interest rates. Once again the volatility associated with the Fed experiment is striking. Volatility is also noticeably higher than average during the 1973–1975 period (the time of the OPEC oil crisis) and immediately after the October 1987 stock market crash. Table 1 reports the unconditional moments of the data. The mean change in interest rates is close to zero, there is significant excess kurtosis, and the correlation

Table 1

Summary statistics relating to weekly first differences in one-month Treasury bill yields reported in annualized percentage terms

The sample period is January 1970 to April 1994, a total of 1,267 observations. The data are plotted in Fig. 3.

Statistic	Estimate	Std. error
Mean	−0.0034	0.0098
Variance	0.1221	0.0180
Skewness	−1.5116	1.2022
Kurtosis	25.3817	10.6057
$Corr(\Delta r_t, r_{t-1})$	−0.0578	0.0279

between Δr_t and r_{t-1} is negative. These properties have all been documented elsewhere.

This section proceeds by estimating several popular models of the short rate noting their strengths and weaknesses. A series of diagnostic tests demonstrates that regime switches are an important feature of the data. This point is made clear by comparing a series of nested models with the GRS model. The goal is to determine what features a model must have in order to capture the mean reversion and conditional heteroskedasticity in short-term interest rates.

All models were estimated by maximum likelihood using the GAUSS MAXLIK and CML modules. The parameter estimates reported were obtained using the Broyden Fletcher, Goldfarb, and Shanno algorithm that is described in Gill, Murray, and Wright (1981), but the results that are reported are robust to the use of the Berndt, Hall, Hall, and Hausman (1974) algorithm, and to different starting values. The standard errors are computed from the diagonal of the heteroskedastic-consistent covariance matrix (see White, 1980).

4.2. The constant-variance model

We begin by considering whether regime-switching alone is sufficient to account for the conditional heteroskedasticity in the short rate. First, we estimate a constant-variance first-order Markov regime-switching model. In this model, changes in the short rate are distributed $N(a_{0i} + a_{1i}r_{t-1}, b_{0i})$ in regime i, so the variance is constant within each regime. We refer to this model as the *constant-variance regime-switching model*. Any conditional heteroskedasticity can only be driven by switches between regimes. The parameter estimates for this model appear in Table 2. The detailed specification of the models appears below the respective tables.

The first column of Table 2 reports maximum likelihood estimates of the single-regime constant-variance model. The conditional mean terms are not significantly different from zero, although a_1 is negative, consistent with some reversion to

Table 2

Parameter estimates and related statistics for single-regime and regime-switching constant-variance models

The sample contains weekly one-month Treasury bill yields reported in annualized percentage terms and extends from January 1970 to April 1994, a total of 1,267 observations.

Parameter	Single regime		Regime switching	
	Estimate	$t\|(p\text{-value})$	Estimate	$t\|(p\text{-value})$
a_{01}	0.0475	1.1757	0.1687	1.2102
a_{02}	–	–	−0.0057	−0.3132
a_{11}	−0.0072	−1.0746	−0.0190	−1.2752
a_{12}	–	–	0.0015	0.4688
b_{01}	0.3489	13.7362*	0.6716	8.5762*
b_{02}	–	–	0.1496	19.6842*
P	–	–	0.9680	53.4807*
Q	–	–	0.9905	202.1429*
Log-likelihood	−463.1433		111.1109	
LB_1^2	56.16	(0.000)	3.25	(0.071)
LB_2^2	74.39	(0.000)	5.23	(0.073)
LB_3^2	78.85	(0.000)	5.71	(0.127)
LB_5^2	96.44	(0.000)	7.46	(0.189)
LB_{10}^2	125.45	(0.000)	10.39	(0.407)
LB_{15}^2	221.39	(0.000)	13.44	(0.569)
$J–B$	123.3865	(0.000)		

t-statistics are based on heteroskedastic-consistent standard errors. Significance of P and Q is relative to 0.5. In the single-regime constant-variance model, the standardized residuals are assumed to have a standard normal distribution. The Jarque–Bera $(J–B)$ statistic tests this. LB_i^2 denotes the Ljung–Box statistic for serial correlation of the squared residuals out to i lags. p-values are in parentheses. In the constant-variance regime-switching model:

$$\Delta r_t|\Phi_{t-1} \sim \begin{cases} N(a_{01} + a_{11}r_{t-1}, b_{01}) & \text{w.p.} \quad p_{1t}, \\ N(a_{02} + a_{12}r_{t-1}, b_{02}) & \text{w.p.} \quad 1 - p_{1t}, \end{cases}$$

$$p_{1t} = (1 - Q)\left[\frac{g_{2t-1}(1 - p_{1t-1})}{g_{1t-1}p_{1t-1} + g_{2t-1}(1 - p_{1t-1})}\right] + P\left[\frac{g_{1t-1}p_{1t-1}}{g_{1t-1}p_{1t-1} + g_{2t-1}(1 - p_{1t-1})}\right],$$

$$g_{1t} = f(\Delta r_t|S_t = 1), \qquad g_{2t} = f(\Delta r_t|S_t = 2).$$

In the single-regime constant-variance model: $\Delta r_t|\Phi_{t-1} \sim N(a_0 + a_1 r_{t-1}, b_0)$.

*Significant at 5%.

the mean. The implied long-run mean $(-a_0/a_1)$ is 6.60% per annum. The table also includes the results of a series of diagnostic tests. Not surprisingly, this constant-variance model does a poor job of modeling the volatility of short-term interest rates. There is serial correlation in the squared standardized residuals,

and the distribution of standardized residuals is skewed and leptokurtic relative to the assumed normal distribution. These results all point towards time-varying conditional (on available information) variances.

The second column of Table 2 reports the results for the constant-variance regime-switching model. In many respects, the results are similar to past work. There is persistence in both regimes with P and Q both exceeding 0.9 and the regimes tend to be separated by differential variances with the regime 1 standard deviation being more than four times the regime 2 standard deviation. Although the conditional mean parameters are all insignificantly different from zero, they do provide an interesting economic result. Regime 1 is characterized as a period of high variance and high interest rates – the implied long-run mean is 8.89% per annum. In this regime, there is weak evidence of reversion to the relatively high long-run mean. When the short rate is 12%, the conditional mean change is -5.93 basis points, which is an economically significant one-week change in interest rates. Moreover, when the short rate is 15%, which is not uncommon during periods of high volatility, the conditional mean change is -11.63 basis points. Alternatively, regime 2 is a low-variance, low-interest rate regime, with a long-run mean of 3.80% per annum, and apparently negative mean reversion. When the short rate is above the long-run mean, it tends to increase further. Since it is rare in our sample for the short rate to be below 3.80%, this acts as a kind of reflecting barrier. However, the strength of this reflection is relatively small: when the short rate is 5% and 8% the expected change is $+0.18$ and $+0.63$ basis points, respectively. In this sense, the short rate essentially follows a random walk during periods of low volatility. This result has important economic implications, as it suggests that the mean reversion, which is assumed by many models to be operating continuously, actually occurs only during occasional periods of high interest rates and high volatility.

The Ljung–Box statistics relating to the squared standardized residuals have been reduced dramatically so that no evidence of serial correlation appears. The simple regime-switching model can capture much of the stochastic volatility of short-term interest rates. This too has clear economic implications. The short rate generally has low volatility, but exhibits very high volatility during some periods. These high volatility periods may have been precipitated by external oil shocks or inflation shocks, for example, and have tended to be persistent. Whereas the regime-switching model is able to incorporate the two different levels of volatility, the single-regime model treats volatility as being constant at an average level. In this case, volatility estimates are uniformly too high during periods of low volatility and uniformly too low during periods of high volatility, hence the highly significant Ljung–Box statistics. Clearly, a different forecast of volatility is required in each regime.

The statistical significance of the second regime cannot be tested using a likelihood ratio test (LRT) because the parameters associated with the second regime are not identified under the null of a single regime. In this case, the LRT statistic

is no longer distributed χ^2 under the null. Hansen (1992) has developed a standardized LRT procedure that overcomes this difficulty. This procedure, however, involves a series of optimizations over a grid of the nuisance parameters, which is extremely burdensome unless the model is very simple or the grid is coarse. In the case at hand, the huge difference in the log-likelihood values corresponds to an LRT statistic of 1,148.5084. While this is not a formal test, because we have not adjusted the χ^2 distribution to take account of the identification problem, it does provide a certain amount of confidence in the existence of a second regime.

In estimating regime-switching models, two different conditional probabilities are of interest. The ex ante probability, $\Pr[S_t = 1|\Phi_{t-1}]$, is of interest in forecasting, based on an evolving information set. This differs from the *filter* probability ($\Pr[S_t = 1|\Phi_t]$) that is commonly reported in the regime-switching literature. The *smoothed* probability, $\Pr[S_t = 1|\Phi_T]$, is of interest in determining if and when regime switches occur. Gray (1995) develops an efficient filter that uses the recursive nature of the regime probabilities and directly links these two conditional probabilities. [2] This efficient recursive representation of the smoothing filter yields the same results as the procedure of Hamilton (1989, 1990) as generalized by Diebold, Lee, and Weinbach (1994). Kim (1994) has developed a relatively efficient algorithm for calculating smoothed probabilities that also makes some use of the recursive nature of the regime probabilities. Kim's algorithm, however, iterates backwards from $\Pr[S_T = 1|\Phi_T]$ to $\Pr[S_{T-1} = 1|\Phi_T]$ and ultimately to $\Pr[S_t = 1|\Phi_T]$. In contrast, the smoothing algorithm derived in Gray (1995) is forward-looking and directly links the ex ante probabilities (of interest in forecasting) with the corresponding smoothed probabilities (of interest to the econometrician).

The top panel of Fig. 4 contains plots of the ex ante and smoothed probabilities. This figure points towards three periods during which the process was in the high-variance regime: 1973–1975, 1979–1983, and late 1987. These periods have quite intuitive explanations in the context of this particular regime-switching model. The first of these periods (1973–1975) corresponds to the OPEC oil crisis, a period known for increased financial market volatility. There is also a marked decline in interest rates corresponding to the recession of 1974. The second period (1979–1983) is clearly driven by the Fed experiment. The unprecedented level of interest rate volatility classifies this as an episode of the high-volatility regime. The third period (1987) is a spike of short duration immediately after the 1987 stock market crash.

The bottom panel of Fig. 4 contains a plot of the conditional standard deviation implied by this simple regime-switching model. The presence of just two

---

[2] This filter also provides a link between the regime-switching literature and the conditional asset pricing literature in the spirit of Campbell (1993) and Ferson and Harvey (1991). The former has concentrated on the ex post identification of regime switches (via smoothed probabilities) while the latter involves modeling the time-varying conditional distribution of asset returns, conditional on available information. This is exactly what the ex ante probability measures.

possible levels of volatility makes the conditional variance mimic the regime probabilities. The GRS model (below) overcomes this limitation through a richer parameterization of the conditional variance within each regime.

4.3. The regime-switching GARCH model

We now relax the assumption of constant variances within each regime, allowing the conditional variances to be GARCH processes of the form:

$$\Delta r_t = a_0 + a_1 r_{t-1} + \varepsilon_t \,,$$

$$\varepsilon_t | \Phi_t \sim N(0, h_t) \,,$$

$$h_t = b_0 + b_1 \varepsilon_{t-1}^2 + b_2 h_{t-1} \,.$$

The first column of Table 3 reports the estimates for the single-regime version of this model. The mean reversion parameter (a_1) is negative as expected, although insignificant, and the implied long-run mean is 4.54%, which is lower than that of the constant-variance model. The conditional variance parameters are similar to those reported elsewhere and are statistically significant (see Engle, Ng, and Rothschild, 1990; Kees, Nissen, Schotman, and Wolff, 1994; Hong, 1988).

As is common in GARCH models of the short rate, $\hat{b}_1 + \hat{b}_2 > 1$, which violates the assumption of stationarity. An LRT is able to reject the hypothesis that the conditional variance follows an IGARCH process where $\hat{b}_1 + \hat{b}_2 = 1$. The LRT statistic is 7.0662 which is χ_1^2 under the null. This issue is dealt with in more detail below. The joint impact of the GARCH effects are statistically important in this single-regime setting. An LRT rejects the single-regime constant-variance model in favor of the single-regime GARCH model. The LRT statistic, which is distributed χ_2^2 under the null, is 1,185.7474, which is significant at any usual level.

The single-regime GARCH model, however, does a poor job of modeling the stochastic volatility in short-term interest rates. The Ljung–Box statistics relating to the squared standardized residuals indicate significant serial correlation, and the Jarque–Bera test indicates that the standardized residuals are not normally distributed (as this model assumes). These results, coupled with the fact that the simple regime-switching model does a better job of accommodating the stochastic volatility, point towards instability of the GARCH parameters. Suppose that a particular GARCH parameter takes a high value in one regime and a low value in the other. A single-regime GARCH model has the effect of averaging this parameter over the sample period so that the model does not do a good job of describing the data in either regime. This can induce positive serial correlation in the standardized squared residuals. This point can be illustrated in the context of a simple ARCH(1) model where $\varepsilon_t = a_0 + a_1 \varepsilon_{t-1}^2 + \eta_t$. Suppose $a_1 = 0.7$ in regime 1 and 0.3 in regime 2, and that in a single-regime model a_1 is

S.F. Gray / Journal of Financial Economics 42 (1996) 27–62

Table 3

Parameter estimates and related statistics for single-regime and regime-switching GARCH models

The sample contains weekly one-month Treasury bill yields reported in annualized percentage terms and extends from January 1970 to April 1994, a total of 1,267 observations.

Parameter	Single regime		Regime switching	
	Estimate	$t\|(p\text{-value})$	Estimate	$t\|(p\text{-value})$
a_{01}	0.0059	0.5017	0.1407	1.1613
a_{02}		–	−0.0011	−0.0762
a_{11}	−0.0013	−0.5160	−0.0141	−1.0807
a_{12}			0.0006	0.2366
b_{01}	0.0004	1.1356	0.1870	2.2278*
b_{11}	0.2095	2.5506*	0.4609	1.9885*
b_{21}	0.8208	13.4603*	0.1977	1.2378
b_{02}	–	–	0.0099	3.8610*
b_{12}	–	–	0.1655	2.7556*
b_{22}	–	–	0.2685	2.8150*
P	–		0.9739	24.1357*
Q	–	–	0.9896	86.8336*
Log-likelihood	129.7304		151.2288	
LB_1^2	6.34	(0.012)	0.22	(0.638)
LB_2^2	6.34	(0.042)	0.26	(0.877)
LB_3^2	11.01	(0.012)	1.66	(0.646)
LB_5^2	13.38	(0.020)	1.75	(0.882)
LB_{10}^2	16.47	(0.087)	2.63	(0.989)
LB_{15}^2	21.57	(0.120)	4.01	(0.998)
$J\text{-}B$	407.1569	(0.000)		

t-statistics are based on heteroskedastic-consistent standard errors. Significance of P and Q is relative to 0.5. In the single-regime GARCH model, the standardized residuals are assumed to have a standard normal distribution. The Jarque–Bera $(J\text{-}B)$ statistic tests this. LB_i^2 denotes the Ljung–Box statistic for serial correlation of the squared residuals out to i lags. p-values are in parentheses. In the regime-switching GARCH model:

$$\Delta r_t | \Phi_{t-1} \sim \begin{cases} N(a_{01} + a_{11}r_{t-1}, h_{1t}) & \text{w.p.} \quad p_{1t}, \\ N(a_{02} + a_{12}r_{t-1}, h_{2t}) & \text{w.p.} \quad 1 - p_{1t}. \end{cases}$$

$$h_{it} = b_{0i} + b_{1i}\varepsilon_{t-1}^2 + b_{2i}h_{t-1},$$

$$\varepsilon_{t-1} = \Delta r_{t-1} - [p_{1t-1}\mu_{1t-1} + (1 - p_{1t-1})\mu_{2t-1}].$$

$$\mu_{it-1} = a_{0i} + a_{1i}r_{t-2}.$$

$$h_{t-1} = p_{1t-1}[\mu_{1t-1}^2 - h_{1t-2}] + (1 - p_{1t-1})[\mu_{2t-1}^2 - h_{2t-2}] - [p_{1t-1}\mu_{1t-1} - (1 - p_{1t-1})\mu_{2t-1}]^2,$$

$$p_{1t} = (1 - Q)\left[\frac{g_{2t-1}(1 - p_{1t-1})}{g_{1t-1}p_{1t-1} + g_{2t-1}(1 - p_{1t-1})}\right] + P\left[\frac{g_{1t-1}p_{1t-1}}{g_{1t-1}p_{1t-1} + g_{2t-1}(1 - p_{1t-1})}\right],$$

$$g_{1t} = f(\Delta r_t | S_t = 1), \qquad g_{2t} = f(\Delta r_t | S_t = 2).$$

In the single-regime GARCH model: $\Delta r_t | \Phi_{t-1} \sim N(a_{01} + a_{11}r_{t-1}, h_{1t})$.

*Significant at 5%.

estimated to be 0.5. Then in regime 1, $\eta_t = 0.2\varepsilon_{t-1}^2 + \text{noise}$, and in regime 2, $\eta_t = -0.2\varepsilon_{t-1}^2 + \text{noise}$. In both cases, η_t exhibits positive serial correlation.

The second column of Table 3 reports estimates of the *regime-switching GARCH model*. The path-independent model developed in Section 4 is used to facilitate tractable estimation of the model. While none of the conditional mean parameters reach statistical significance, they confirm the asymmetry across regimes evident in the constant-variance regime-switching model. In the high volatility regime (regime 1) the implied long-run mean is 9.98%, so that when the short rate is 15% the conditional mean change is −7.08 basis points, for example. In the low-volatility regime (regime 2) the implied long-run mean is 1.83%, which again plays the role of a reflecting barrier as there appears to be negative mean reversion. In the constant-variance regime-switching model, however, the short rate essentially follows a random walk during periods of low volatility. When the short rate is 8%, for example, the conditional mean change is only + 0.37 basis points. Consistent with the results for the constant-variance
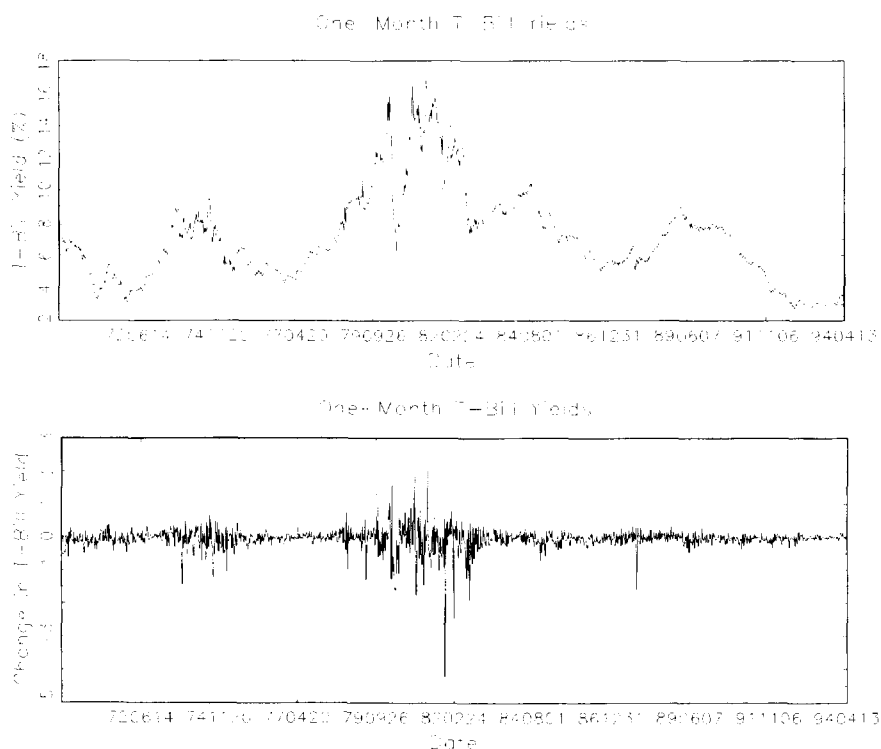

Fig. 3. The top panel contains a time series plot of weekly one-month Treasury bill rates, reported in annualized percentage terms. The first differences of this series are shown in the bottom panel. The sample period is January 1970 to April 1994, a total of 1,267 observations.
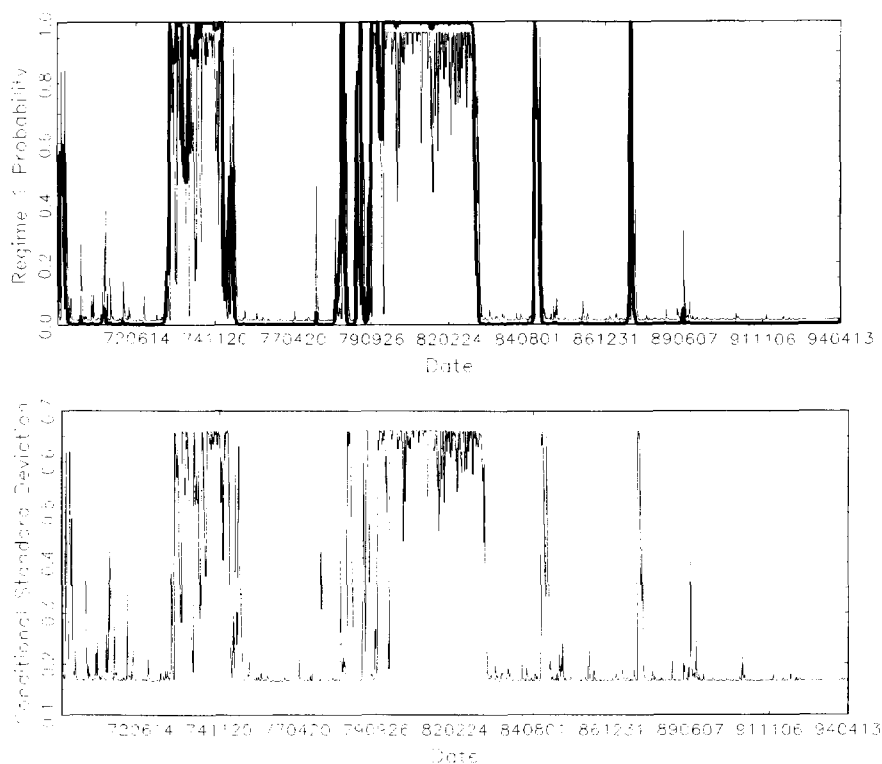
Fig. 4. The top panel contains a time series plot of the ex ante (thin line) and smoothed (bold line) probabilities that the short rate process is in regime 1 (the high-volatility regime) at time t according to the constant-variance regime-switching model. The ex ante probability is based on information available at time t ($\Pr[S_t = 1|\Phi_{t-1}]$) and the smoothed probability is based on the entire sample ($\Pr[S_t = 1|\Phi_T]$). The bottom panel contains a time series plot of the conditional standard deviation of changes to the short rate based on the constant-variance regime-switching model ($\sqrt{\mathrm{var}[\Delta r_t|\Phi_{t-1}]}$). Parameter estimates are based on a data set of weekly one-month Treasury Bill rates, reported in annualized percentage terms. The sample period is January 1970 to April 1994, a total of 1,267 observations.

regime-switching model, there is some evidence of mean reversion at high levels of the interest rate (coinciding with episodes of regime 1 and high volatility), but there is no evidence whatsoever of mean reversion at low and moderate interest rates (coinciding with episodes of regime 2 and lower volatility).

Further, the high-volatility regime is characterized by more sensitivity to recent shocks ($b_{11} > b_{12}$) and less persistence ($b_{21} < b_{22}$) than the low-volatility regime. The effect of individual shocks dies out quickly during periods of very high volatility but has a longer-lasting effect during periods of low to moderate volatility. It is this important difference that a single-regime GARCH model is unable to capture. For example, in the single-regime GARCH model, a shock of
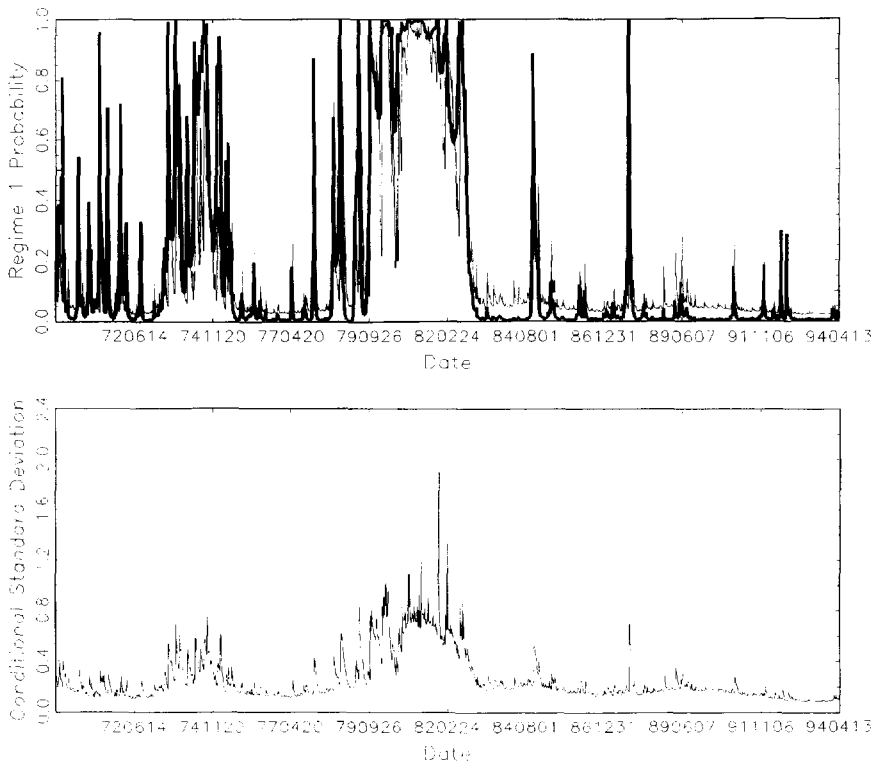
Fig. 5. The top panel contains a time series plot of the ex ante (thin line) and smoothed (bold line) probabilities that the short rate process is in regime 1 (the high-volatility regime) at time t according to the GRS model. The ex ante probability is based on information available at time t ($\Pr[S_t = 1 | \Phi_{t-1}]$) and the smoothed probability is based on the entire sample ($\Pr[S_t = 1 | \Phi_T]$). The bottom panel contains a time series plot of the conditional standard deviation of changes to the short rate based on the GRS model ($\sqrt{\mathrm{var}[\Delta r_t | \Phi_{t-1}]}$). Parameter estimates are based on a data set of weekly one-month Treasury bill rates, reported in annualized percentage terms. The sample period is January 1970 to April 1994, a total of 1267 observations.

1% to the short rate at time t increases the conditional variance at times $t + 1$ to $t + 5$ by 0.2095, 0.1720, 0.1411, 0.1159, and 0.0951, respectively. In the regime-switching model, the corresponding increases to the conditional variance in the high-volatility regime are 0.4609, 0.0911, 0.0180, 0.0036, and 0.0007. The corresponding increases to the conditional variance in the low-volatility regime are 0.1655, 0.0444, 0.0119, 0.0032, and 0.0009. Note that shocks appear to be much more persistent in the single-regime model. In the regime-switching model, during periods of high volatility, shocks have a large immediate impact that dies out very quickly. Friedman and Laibson (1989) present a modified ARCH (MARCH) model of stock returns that is consistent with these results in several respects.

In their model, the persistence of individual shocks differs according to the magnitude of the shock. They model small to moderate shocks as being persistent whereas very large shocks (extraordinary volatility) are nonpersistent. Once again, these results seem economically reasonable in terms of the 'pressure-relieving' argument outlined above. Moreover, volatility in the short rate can be the result of uncertainty about the future direction of the economy and inflation expectations. In such cases, it is not uncommon for the monetary authority to signal its intention by exerting a strong force on short-term interest rates. The result can be a large change in the short rate and a reduction in uncertainty surrounding inflation expectations which causes a large nonpersistent shock.

Within each regime, the GARCH processes are stationary ($b_{1i} + b_{2i} < 1$) and *much* less persistent than in the single-regime GARCH model, illustrating another potential advantage of the regime-switching model over the single-regime GARCH model. There are really two sources of volatility persistence. If one regime has low average variance and one regime has high average variance, and if the regimes are persistent, volatility will be persistent. A long period of low average variance may be followed by a long period of high average variance if the regime switches and then persists. The parameter estimates here suggest that the regimes are very persistent, so this source of volatility persistence will be important. Additionally, if the effect of an individual shock takes a long time to die out, reverting to the average variance in that regime, there is within-regime persistence. The parameter estimates reported here indicate that (1) the regimes are very persistent and (2) within-regime persistence is much lower than in the single-regime GARCH model. Since the single-regime GARCH model cannot capture the persistence of regimes, all of the persistence in volatility is thrown into the persistence of an individual shock. Individual shocks then appear to take too long to die down to the average variance. In the regime-switching model, the effect of individual shocks dies down quickly to the average variance of the particular regime (especially so for the high-volatility regime). This is consistent with the findings of Lamoureux and Lastrapes (1990); the untenable results produced in some GARCH models may be caused by trying to use a single-regime model to capture a multi-regime process.

To test the significance of the GARCH parameters, an LRT was constructed to compare the constant-variance regime-switching model (column 2 of Table 2) with the GARCH regime-switching model. The LRT statistic, which is distributed χ_4^2 under the null, is 80.2358, which is significant at any usual level, indicating that the GARCH effects are important. Once again, it is difficult to test the statistical significance of the second regime. When comparing the regime-switching GARCH model with the single-regime GARCH model, there are five parameters that are unidentified under the null of a single regime (when $P = 1$ and $Q = 0$). In this case, the quasi-LRT statistic, which is 42.9968 and highly significant under ordinary circumstances, can no longer be assumed to be distributed as a χ^2. We therefore also appeal to the economic significance of the second regime. That is,

the behavior of the short rate is substantially different in each regime. During periods of high and volatile interest rates (regime 1) there is some evidence of reversion to a mean of around 10%, and individual shocks have a large immediate effect on the conditional variance, but this effect dies out quickly. During periods of low and stable interest rates (regime 2) the short rate follows a low-variance random walk and individual shocks have a smaller immediate effect on the conditional variance, but this effect is relatively more persistent. To the extent that these two regimes produce substantially different forecasts of the short rate and its volatility, we consider the addition of a second regime to be economically significant. We reexamine this issue in more detail in Section 5.

In plots of the ex ante and smoothed probabilities (not shown here) the periods of high volatility associated with the Fed experiment, OPEC oil crises, and stock market crash are again apparent. Further, the regime-switching GARCH model does a relatively good job of modeling the stochastic volatility in short-term interest rates. The Ljung–Box statistics relating to the squared standardized residuals indicate no remaining serial correlation. Also, the extra flexibility of allowing GARCH effects in each regime provides a richer characterization of the conditional variance. No longer do conditional variance estimates mimic regime probabilities. The volatility associated with the Fed experiment is clearly greater than that associated with other high-volatility periods, as expected.

4.4. The generalized regime-switching model

In the final group of models, the conditional variance process is augmented to accommodate level effects as well as GARCH effects. In these models, the constant term in the GARCH equation (b_{0i}) is omitted. This is because a de facto intercept is introduced by the square root term. That is, $\sigma^2 \min(r_{t-1})$ forms a lower bound for the conditional variance. In experimenting with unconstrained models, the (b_{0i}) term always converges to the boundary ($b_{0i} = 0$) and was therefore dropped from the model. Estimates of the single-regime version of this model appear in the first column of Table 4. Again, the mean reversion parameter is insignificant. The level of interest rates enters significantly into the conditional variance function. Both GARCH and level effects appear to be important in characterizing conditional variances. Two factors appear to be important in determining volatility: recent volatility and high interest rates. The diagnostic tests indicate that this model suffers from the same problems as the single-regime GARCH model.

Estimates of the regime-switching version of this model appear in the second column of Table 4. The estimates of the conditional mean parameters are similar in size and interpretation to those of the regime-switching GARCH model. Once again, there is some reversion to a relatively high long-run mean during periods of high interest rates and high volatility, whereas the short rate follows a low-variance random walk during periods of low and stable interest rates.

Table 4

Parameter estimates and related statistics for single-regime and regime-switching GRS models

The sample contains weekly one-month Treasury bill yields reported in annualized percentage terms and extends from January 1970 to April 1994, a total of 1,267 observations.

Parameter	Single regime		Constant probabilities		Time-varying probabilities	
	Estimate	$t\|(p\text{-value})$	Estimate	$t\|(p\text{-value})$	Estimate	$t\|(p\text{-value})$
a_{01}	0.0046	0.4072	0.0550	0.5848	0.0401	0.2987
a_{02}	–	–	0.0014	0.1058	−0.0009	−0.0706
a_{11}	−0.0008	−0.3445	−0.0060	−0.5512	−0.0063	−0.4596
a_{12}	–	–	0.0002	0.0932	0.0009	0.3925
b_{11}	0.2269	2.5794*	0.3038	1.7434	0.2390	1.2107
b_{21}	0.7893	11.1508*	0.0936	0.6882	0.0987	−0.3510
σ_1	0.0133	3.3169*	0.1550	6.1633*	0.1726	2.4996*
b_{12}	–	–	0.1528	2.7021*	0.1744	3.3731*
b_{22}	–	–	0.2383	2.3614*	0.3474	0.6385
σ_2	–	–	0.0420	8.1531*	0.0327	0.7940
P/c_1	–	–	0.9688	47.3653*	−0.4384	−0.1173
d_1	–	–	–	–	0.2011	0.6079
Q/c_2	–	–	0.9881	143.1565*	2.1989	1.3827
d_2	–	–	–	–	−0.0421	−0.2388
Log-likelihood	136.0330		176.4741		182.6711	
LB_1^2	4.95	(0.026)	0.04	(0.851)	0.02	(0.903)
LB_2^2	4.96	(0.084)	0.29	(0.867)	0.04	(0.978)
LB_3^2	9.67	(0.022)	1.95	(0.583)	2.11	(0.550)
LB_5^2	11.86	(0.037)	1.96	(0.854)	2.26	(0.813)
LB_{10}^2	14.11	(0.168)	2.79	(0.986)	7.11	(0.715)
LB_{15}^2	18.52	(0.236)	5.02	(0.992)	11.65	(0.705)
$J–B$	400.50	(0.000)				

t-statistics are based on heteroskedastic-consistent standard errors. Significance of P and Q is relative to 0.5. In the single-regime GRS model, the standardized residuals are assumed to have a standard normal distribution. The Jarque–Bera $(J–B)$ statistic tests this. LB_i^2 denotes the Ljung–Box statistic for serial correlation of the squared residuals out to i lags. p-values are in parentheses. In the full GRS model:

$$\Delta r_t|\Phi_{t-1} \sim \begin{cases} N(a_{01} + a_{11}r_{t-1}, v_{1t}) & \text{w.p.} \quad p_{1t}, \\ N(a_{02} + a_{12}r_{t-1}, v_{2t}) & \text{w.p.} \quad 1 - p_{1t}, \end{cases}$$

$$v_{it} = h_{it} + \sigma_i^2 r_{t-1},$$

$$h_{it} = b_{1i}\varepsilon_{t-1}^2 + b_{2i}v_{t-1},$$

$$\varepsilon_{t-1} = \Delta r_{t-1} - [p_{1t-1}\mu_{1t-1} + (1 - p_{1t-1})\mu_{2t-1}],$$

$$\mu_{it-1} = a_{0i} + a_{1i}r_{t-2},$$

$$v_{t-1} = p_{1t-1}[\mu_{1t-1}^2 + v_{1t-1}] + (1 - p_{1t-1})[\mu_{2t-1}^2 + v_{2t-1}] - [p_{1t-1}\mu_{1t-1} + (1 - p_{1t-1})\mu_{2t-1}]^2,$$

$$p_{1t} = (1 - Q_t)\left[\frac{g_{2t-1}(1 - p_{t-1})}{g_{1t-1}p_{t-1} + g_{2t-1}(1 - p_{t-1})}\right] + P_t\left[\frac{g_{1t-1}p_{t-1}}{g_{1t-1}p_{t-1} + g_{2t-1}(1 - p_{t-1})}\right],$$

$$g_{1t} = f(\Delta r_t|S_t = 1), \qquad g_{2t} = f(\Delta r_t|S_t = 2),$$

$$P_t = \Phi(c_1 + d_1r_{t-1}), \qquad Q_t = \Phi(c_2 + d_2r_{t-1}).$$

In the GRS model with constant switching probabilities $P_t = P$ and $Q_t = Q$. In the single-regime GRS model: $\Delta r_t|\Phi_{t-1} \sim N(a_{01} + a_{11}r_{t-1}, v_{1t})$.

*Significant at 5%.

The conditional variances appear to separate into a GARCH regime and a CIR regime. In regime 1, both GARCH parameters fail to reach significance, while the CIR parameter is highly significant. In regime 2, both GARCH parameters are significant and although the CIR parameter is statistically significant its small value renders it economically insignificant. When the interest rate is 5%, which is about the mean in the low-volatility regime, $\hat{\sigma}^2 r_{t-1} = 0.009$, while the average level of the conditional variance is in the order of 0.12.

The results indicate that the persistence of an individual shock is much less than is suggested by the single-regime model, with $\hat{b}_{1i} + \hat{b}_{2i} < 0.4$ in both regimes, while in the single-regime model $\hat{b}_1 + \hat{b}_2 > 1$. Conversely, allowing for regime switches has increased the value of the CIR parameter in both regimes. The economic cause of these results is that in the single-regime model the only source of clustering in volatility comes through the GARCH process. In the regime-switching model, volatility clustering can be caused by three factors. First, the GARCH process in each regime is clearly capable of capturing volatility clustering; indeed this is the whole point of GARCH models. Second, if the unconditional variance (or average level of conditional variance) is higher in one regime than the other, and if regimes are somewhat persistent, then periods of high volatility will cluster together during episodes of the high-volatility regime. Third, since volatility depends on the level of interest rates, volatility clustering can result during periods of high interest rates, if interest rates are persistent. Again note that during periods of very high volatility, the conditional variance is more sensitive to recent shocks, but the effect of these shocks dies out relatively quickly.

Allowing the transition probabilities to be state-dependent (this is the full GRS model) significantly improves the performance of the model. An LRT clearly rejects the constant probability model (column 2 of Table 4) in favor of the full GRS model (column 3 of Table 4). The LRT statistic, which is distributed χ^2_2 under the null, is 12.3940 which is significant at the 1% level. In interpreting the c_i and d_i coefficients, first note that c_2 and d_2 affect Q_t, the probability of remaining in regime 2, the random-walk/low-variance regime. Since d_2 is negative, the probability of staying in this regime decreases as the level of interest rates increases. Conversely, since d_1 is positive, the probability of staying in the high-variance /mean-reversion regime decreases as the level of interest rates falls. Therefore, when interest rates increase, the probability of staying in or switching to the mean-reverting regime increases. This helps to prevent the interest rate process from 'wandering off' into unreasonable regions. Under the random walk regime, interest rates are unbounded. However, as interest rates increase, the process is more likely to switch to the regime in which interest rates tend to revert to a long-run mean. Conversely, when interest rates are low, the random-walk/low-volatility regime is more likely.

The GRS model does a relatively good job of modeling the stochastic volatility in short-term interest rates. The Ljung–Box statistics relating the squared

standardized residuals indicate no serial correlation. The volatility of interest rates behaves as follows. During 'normal' periods of low to moderate volatility, the most important factor in characterizing the conditional variance is volatility clustering, and hence the conditional variance process is well described by a GARCH model. During periods of extreme volatility, however, the persistence in volatility (associated with GARCH models) decreases dramatically and the conditional variance increases with the square root of the level of the interest rate.

To establish the importance of allowing for a second regime, we again appeal to its economic significance, due to the problems in identifying nuisance parameters under the null of a single regime. In this case, the quasi-LRT statistic is 93.2762 which would be highly significant under ordinary circumstances, although this statistic can no longer be assumed to be distributed as a χ^2. The behavior of the short rate is substantially different in each regime, confirming the results for the regime-switching GARCH model. During periods of high and volatile interest rates (regime 1) there is some evidence of reversion to a relatively high mean, the conditional variance is related to the level of the short rate, and the effect of individual shocks dies out quickly. During periods of low and stable interest rates (regime 2) the short rate follows a low-variance random walk and individual shocks have a smaller, but more persistent, effect on the conditional variance. Once again, to the extent that these two regimes produce substantially different forecasts of the short rate and its volatility, we consider the addition of a second regime to be economically significant.

The top panel of Fig. 5 contains plots of the ex ante and smoothed probabilities from this model. Again, the Fed experiment and OPEC oil crisis are particularly apparent. The bottom panel of Fig. 5 plots the conditional standard deviation implied by this model. Table 5 contains a summary of the various models that have been estimated.

5. Economic implications of regime-switching

In this section we seek to further establish the economic importance of allowing for switches between different regimes. By characterizing the behavior of the short rate as being substantially different across regimes, we demonstrate that the GRS model performs well in forecasting volatility out-of-sample, and show how the GRS model can yield security prices that are different from those produced by existing models.

5.1. A characterization of two regimes

The GRS model identifies a number of features of short-term interest rates that existing models are unable to capture. In particular, the short-term interest rate can be characterized as having two distinct regimes, spending the majority

Table 5

Summary of restrictions placed on models that are nested within the GRS framework (nuisance parameters are those that are not identified under the null of a single regime)

Model	Restrictions	Nuisance parameters
(1) Single-regime constant-variance model	$d_1 = d_2 = 0$ $c_1 = \infty,\ c_2 = 0$ $b_{11} = b_{21} = \sigma_1 = 0$	$a_{02},\ a_{12}$ $b_{02},\ b_{12},\ b_{22},\ \sigma_2$
(2) Regime-switching constant-variance model	$d_1 = d_2 = 0$ $b_{11} = b_{21} = \sigma_1 = 0$ $b_{12} = b_{22} = \sigma_2 = 0$	
(3) Single-regime GARCH model	$d_1 = d_2 = 0$ $c_1 = \infty,\ c_2 = 0$ $\sigma_1 = 0$	$a_{02},\ a_{12}$ $b_{02},\ b_{12},\ b_{22},\ \sigma_2$
(4) Regime-switching GARCH model	$d_1 = d_2 = 0$ $\sigma_1 = \sigma_2 = 0$	
(5) Single-regime GRS model	$d_1 = d_2 = 0$ $c_1 = \infty,\ c_2 = 0$ $b_{01} = 0$	$a_{02},\ a_{12}$ $b_{02},\ b_{12},\ b_{22},\ \sigma_2$
(6) Regime-switching GRS model	$d_1 = d_2 = 0$ $b_{01} = b_{02} = 0$	
(7) Full GRS model	$b_{01} = b_{02} = 0$	

The most general form of the model is:

$$\Delta r_t | \Phi_{t-1} \sim \begin{cases} N(a_{01} + a_{11}r_{t-1}, v_{1t}) & \text{w.p.} \quad p_{1t}. \\ N(a_{02} + a_{12}r_{t-1}, v_{2t}) & \text{w.p.} \quad 1 - p_{1t}. \end{cases}$$

$$v_{it} = h_{it} + \sigma_i^2 r_{t-1}.$$

$$h_{it} = b_{0i} + b_{1i}\varepsilon_{t-1}^2 + b_{2i}v_{t-1}.$$

$$\varepsilon_{t-1} = \Delta r_{t-1} - [p_{1t-1}\mu_{1t-1} + (1 - p_{1t-1})\mu_{2t-1}],$$

$$\mu_{it-1} = a_{0i} + a_{1i}r_{t-2}.$$

$$v_{t-1} = p_{1t-1}[\mu_{1t-1}^2 + v_{1t-1}] + (1 - p_{1t-1})[\mu_{2t-1}^2 + v_{2t-1}] - [p_{1t-1}\mu_{1t-1} + (1 - p_{1t-1})\mu_{2t-1}]^2.$$

$$p_{1t} = (1 - Q_t)\left[\frac{g_{2t-1}p_{t-1}}{g_{1t-1}p_{t-1} + g_{2t-1}(1 - p_{t-1})}\right] - P_t\left[\frac{g_{1t-1}p_{t-1}}{g_{1t-1}p_{t-1} + g_{2t-1}(1 - p_{t-1})}\right].$$

$$g_{1t} = f(\Delta r_t | S_t = 1), \qquad g_{2t} = f(\Delta r_t | S_t = 2).$$

$$P_t = \Phi(c_1 + d_1 r_{t-1}), \qquad Q_t = \Phi(c_2 + d_2 r_{t-1}).$$

of the time in the low-volatility/random-walk regime with occasional switches to the high-volatility/mean-reverting regime. The conditional variance in the low-volatility regime appears to follow a GARCH (1,1) process with substantially less persistence than is usually reported for short-term interest rates. The reduction in persistence is even more dramatic than that achieved by the Markov ARCH models of Cai (1994) and Hamilton and Susmel (1994). The conditional variance in the high-volatility regime takes a different form with the level of interest

rates becoming an economically and statistically significant predictor of volatility. While the overall GARCH effects are much less important, the sensitivity to recent shocks is greater. The effect of individual shocks is greater (initially) during periods of very high volatility, but has a longer lasting effect during periods of low to moderate volatility. Allowing the transition probabilities to be state-dependent is also statistically and economically significant as the level of interest rates is an important predictor of switches between regimes. As interest rates increase, the likelihood of the process being in the high-volatility/mean-reverting regime increases. The low-volatility regime is technically nonstationary, being essentially a random walk with very small positive drift. In a large sample there is a positive probability of the short rate drifting upwards without bound, although the conditional variance is so low in this regime that the sample would have to be *very* large. As the short rate increases, however, the probability of switching into the other regime (with strong mean reversion) increases. This prevents the short rate from 'drifting off' into unreasonable regions.

5.2. Out-of-sample specification tests

To avert fears of overparameterization of the GRS model, and to establish the economic significance of allowing for a second regime, we report the results of a series of out-of-sample tests in Table 6. In performing these tests, we estimate the parameters of the particular model over one period and compute a time series of conditional variances over a subsequent period, holding the parameters fixed. Since the conditional variance is an expectation of squared innovations to the interest rate process, we compare this to the actual squared innovations. The difference between forecast and actual squared innovations, over in-sample and out-of-sample periods, is then summarized in the form of root mean squared errors (RMSE), mean absolute errors (MAE), and the R^2 between actual volatility ($av_t = \varepsilon_t^2$ where $\varepsilon_t = \Delta r_t - E_{t-1}[\Delta r_t]$) and forecast volatility ($fv_t = E_{t-1}[\varepsilon_t^2]$). In particular,

$$R^2 = 1 - \frac{\sum_{t=1}^{T}(av_t - fv_t)^2}{\sum_{t=1}^{T} av_t^2} .$$

This measure of R^2, in providing a direct measure of the goodness of fit of the forecast, differs from the R^2 that would be obtained by projecting actual volatility on forecast volatility. It imposes an intercept of zero and a slope of one on such a projection, permitting direct conclusions about a particular forecast rather than about some linear transformation of that forecast.

In the first test, the models are estimated from the start of the sample up to the Fed experiment, a period that includes the oil shock. The out-of-sample period includes the Fed experiment, the 1987 stock market crash, and the subsequent

Table 6
Out-of-sample specification tests

The table reports root mean squared prediction errors of the difference between actual volatility v_t^2, where $v_t = \Delta r_t - E_{t-1}[\Delta r_t]$, and the conditional variance, which is $E_{t-1}[v_t^2]$. Parameters are estimated over the in-sample period, and held fixed over the out-of-sample period. The constant variance model is the single-regime model described in the first column of Table 2. The GARCH model is the single-regime model described in the first column of Table 3. The GRS model is the full regime-switching model with time-varying switching probabilities described in the final column of Table 4. The data are weekly observations of one-month Treasury bill yields reported in annualized percentage terms. The sample extends from January 1970 to April 1994, a total of 1,267 observations.

| In sample | | | | | | Out of sample | | | | | |
Period (No. obs.)	Statistic	Constant variance	GARCH	GRS		Period (No. obs.)	Statistic	Constant variance	GARCH	GRS
January 1970– December 1978 (470)	RMSE	0.1847	0.1857	0.1835		January 1979– December 1994 (796)	RMSE	0.7760	0.7895	0.7625
	MAE	0.0805	0.0787	0.0711			MAE	0.1704	0.1990	0.1778
	R^2	0.1136	0.1253	0.1439			R^2	0.0261	0.0352	0.0862
January 1970– December 1981 (626)	RMSE	0.8362	0.8682	0.8127		January 1982– April 1994 (640)	RMSE	0.3332	0.3294	0.3172
	MAE	0.2623	0.2554	0.2215			MAE	0.1885	0.0859	0.1244
	R^2	0.0486	0.0003	0.1115			R^2	-0.1511	-0.0529	0.0471
January 1976– December 1982 (365)	RMSE	1.1233	1.2080	0.5528		January 1983– April 1994 (488)	RMSE	0.3041	0.1414	0.1363
	MAE	0.4264	0.4396	0.3737			MAE	0.2687	0.0602	0.0583
	R^2	0.0709	-0.0212	0.1279			R^2	-4.3375	-0.1311	0.0419
January 1970– April 1991 (1110)	RMSE	0.6686	0.6870	0.6378		April 1991– April 1994 (156)	RMSE	0.1287	0.0184	0.0178
	MAE	0.1924	0.1807	0.1561			MAE	0.1274	0.0137	0.0140
	R^2	0.0405	0.0098	0.1351			R^2	-36.9382	-0.0053	0.0370

Negative R^2 values result when incorporating the forecast results in an unexplained sum of squares (the square of the difference between actual and forecast volatility) that is *higher* than the original total sum of squares (the square of the actual volatility). This occurs, for example, when the estimation period has much higher volatility than actually occurs over the out-of-sample period. In this case, the squared difference between forecast and actual volatility can be very large relative to squared actual volatility over the out-of-sample period.

period of lower volatility through to the end of the sample. The GRS model performs well in this experiment, alleviating fears of overparameterization. The parameters of the high-volatility regime are well-identified by the oil shock, so that the GRS model outperforms the other models over the holdout sample, which contains periods of very high and very low volatility.

For the second test, the models are estimated over the first half of the sample and forecasts are made over the second half of the sample. The in-sample estimation period, which includes the oil shock and part of the Fed experiment, is more volatile than the holdout period, which contains the remainder of the Fed experiment and the stock market crash. The constant-volatility model therefore predicts high levels of volatility out-of-sample, whereas interest rates turn out to be relatively more stable. This results in high forecast errors and a negative R^2; the forecast is so high relative to the actual volatility that the difference between forecast and actual volatility exceeds the actual volatility itself. The GARCH model also produces a negative R^2 for the same reason. Once again, the GRS model performs well both in-sample and out-of-sample.

In the third test, the in-sample period includes the Fed experiment and the preceding three years, and the out-of-sample period consists of the period subsequent to the Fed experiment. This test is designed to show that the parameters of the GRS model can be identified by a single episode of each regime as the estimation period contains a period of low volatility and the Fed experiment. In this case, the volatility of the in-sample period is much greater than that of the holdout period, which causes the poor performance of the constant-variance model.

The final test examines the short-term forecasting ability of the model in estimating the models over the entire sample except for the last three years. The in-sample estimation period (which contains the oil shock and the Fed experiment) is, once again, much more volatile than the holdout period. The variance of short rate changes is 17 times larger in the in-sample period than in the holdout period. The constant-volatility model therefore predicts extremely high levels of volatility out-of-sample, whereas interest rates turn out to be very stable. This once again results in high forecast errors and a negative R^2. The GARCH and GRS models perform better, although the GARCH model produces a negative R^2 for the same reason.

In summary, the GRS model performs well in all four tests over three different measures. We conclude that (1) the description of the short-rate process as switching between two different regimes is not due to overfitting, and (2) volatility forecasts from the GRS model may be profitably employed in the valuation of interest-rate-sensitive securities and interest rate risk management.

5.3. Practical uses

The GRS model has a number of important practical applications. One of the recommendations made by the Group of Thirty's Global Derivative Study Group

in *Derivatives: Practices and Principles*, for example, is that derivatives dealers and end-users compute 'value at risk' on a daily basis. Value at risk is defined as the range of market values that a position can attain over a short period of time and is usually determined by setting a 95% confidence interval around the current position value. To compute the value at risk of a book of interest rate securities, the GRS model can provide the necessary forecasts of the mean and the variance of the short rate.

Second, many complex interest-rate-sensitive securities are valued by Monte Carlo simulation under the assumption of risk neutrality. For example, it is common in practice to value complex mortgage-backed securities by simulating a series of interest rates that feeds into a prepayment model that determines the sequence of cash flows that the security might generate. To the extent that the simulated series is more reasonable (in a sense to be made precise below), the performance of the prepayment model, and the pricing model as a whole, will be improved.

Consider, for example, simulating a time series of interest rates to value a short-dated security using Monte Carlo methods. Assume that the GRS model has been estimated using data up to the present time and that the results indicate a high probability that the process is currently in the low-volatility regime. In this case, the simulation would be started with a high probability of drawing the first observation from the low-volatility regime. Because the regimes are quite persistent, this simulated series would look quite different from a series which began with a high probability of drawing from the high-volatility regime.

The economic implication of these results is that existing single-regime models, in effect, average the two regimes and hence produce parameter estimates that are unsatisfactory in modeling either regime. A single-regime model, for example, predicts the same moderate amount of mean reversion to the same long-run mean at all times. The GRS model, however, shows that when interest rates are high (and volatile) strong mean reversion exists, while when interest rates are moderate, the short rate follows a random walk. Series simulated using a single-regime model will look quite different from the process simulated from the GRS model. One way to avoid this averaging effect is to use only recent data in estimating the single-regime model, so that it is more likely that all of the data come from the same regime. However, this approach completely discounts the possibility of a regime switch in the future, which is (historically) nonzero. If the model is estimated during an episode of the low-volatility regime, the possibility of an episode of the high-volatility regime is ignored. This may result in undervaluation of interest-rate-sensitive options, or overvaluation of some barrier options such as kickouts (or down and out or up and out options) as the possibility of larger movements in interest rates is discounted. Conversations with investment bankers confirm that this type of apparent mispricing is quite common in the market for exotic derivatives.

Third, the forecast volatility can be used within the standard log-normal option pricing framework. Whereas this framework assumes that changes (returns) are normally distributed, regime-switching models are based on a mixture of normal distributions. In the results reported in this paper, however, none of the conditional mean parameters are statistically significant, so that the mean change in interest rates is close to zero in both regimes. In this case, the mixture of normal distributions is unimodal and can be approximated by a single normal distribution. This can be contrasted with the case when the two means are significantly different, in which case the mixture can be bimodal.[3] Alternatively, this idea could be formalized, while maintaining the intuition for the GRS conditional variance process. To do this we set

$$\ln(r_t) - \ln(r_{t-1}) \sim N(0, h_t),$$

with $h_t = p_{1t}h_{1t} + (1 - p_{1t})h_{2t}$, where $h_{it} = \text{var}[\ln(r_t) - \ln(r_{t-1})|\Phi_{t-1}, S_t = i]$ as in Eq. (9).

6. Conclusions

This paper develops a generalized regime-switching (GRS) model of the short-term interest rate. The short rate exhibits a different degree of mean reversion and a different form of conditional heteroskedasticity in each regime. The form of the conditional variance in each regime is general and nests the popular GARCH and CIR (square root process) specifications. A first-order Markov process with state-dependent transition probabilities governs the switching between regimes. The empirical results indicate that all of these generalizations are statistically and economically significant.

While the GRS model outperforms simple single-regime models in an out-of-sample forecasting experiment, further research into the practical applications of this kind of model seems warranted. In particular, further attention could be devoted to the security valuation and risk management applications outlined above. These and other related issues are currently being explored in ongoing research.

Appendix: Construction of log-likelihood function

To see how the likelihood function for the GRS model is constructed, first note that when Δr_t is subject to switches between two regimes:

---

[3] I am grateful to the referee for making this point.

$$f(\Delta r_t | \Phi_{t-1}) = \sum_{i=1}^{2} f(\Delta r_t, S_t = i | \Phi_{t-1})$$

$$= \sum_{i=1}^{2} f(\Delta r_t | S_t = i, \Phi_{t-1}) \Pr(S_t = i | \Phi_{t-1})$$

$$= \sum_{i=1}^{2} f(\Delta r_t | S_t = i, \Phi_{t-1}) p_{it},$$

where p_{it} is $\Pr(S_t = i | \Phi_{t-1})$. Therefore, the distribution of Δr_t conditional on available information may be written

$$\Delta r_t | \Phi_{t-1} \sim \begin{cases} f(\Delta r_t | S_t = 1, \ \Phi_{t-1}) & \text{w.p.} \quad p_{1t}, \\ f(\Delta r_t | S_t = 2, \ \Phi_{t-1}) & \text{w.p.} \quad p_{2t}. \end{cases} \tag{A.1}$$

Assuming conditional normality (conditional on available information and on the regime) yields

$$f(\Delta r_t | S_t = i, \Phi_{t-1}) = \frac{1}{\sqrt{2\pi h_{it}}} \exp \left\{ \frac{-(\Delta r_t - \mu_{it})^2}{2 h_{it}} \right\}. \tag{A.2}$$

Finally, note that the regime probabilities of the Hamilton first-order Markov model can be reformulated as a simple recursive process by rewriting the Hamilton model in terms of the recursive structure of $p_{1t} = \Pr(S_t = 1 | \Phi_{t-1})$, that is, the probability that the process is in regime 1 at time t conditional on available information. Thus, the focus is now on the regime probability (the probability of being in a particular regime) rather than the switching probability (the probability of switching from one particular regime to another). If the only information used as conditioning information for p_{1t} is lagged values of the dependent variable (as is the case in this paper), $p_{1t} = \Pr(S_t = 1 | \tilde{r}_{t-1})$ where $\tilde{r}_{t-1} = \{r_{t-1}, r_{t-2}, \ldots\}$. A proof of the more general case can be found in Gray (1995) and Hamilton (1994).

According to the first-order Markov structure, $p_{1t} = \Pr(S_t = 1 | \tilde{r}_{t-1})$ depends only on the regime the process is in at time $t - 1$. By conditioning on the regime at time $t - 1$:

$$\Pr(S_t = 1 | \tilde{r}_{t-1}) = \sum_{i=1}^{2} \Pr(S_t = 1 | S_{t-1} = i, \tilde{r}_{t-1}) \Pr(S_{t-1} = i | \tilde{r}_{t-1}). \tag{A.3}$$

In the Hamilton framework the switching probabilities have a Markov structure such that

$$\begin{aligned} \Pr[S_t = 1 | S_{t-1} = 1] &= P, \\ \Pr[S_t = 2 | S_{t-1} = 1] &= (1 - P), \\ \Pr[S_t = 2 | S_{t-1} = 2] &= Q, \\ \Pr[S_t = 1 | S_{t-1} = 2] &= (1 - Q), \end{aligned} \tag{A.4}$$

in which case

$$\Pr(S_t = 1 | S_{t-1} = i, \tilde{r}_{t-1}) = \Pr(S_t = 1 | S_{t-1} = i). \tag{A.5}$$

Substituting (A.5) and (A.4) into (A.3) yields

$$\Pr(S_t = 1 | \tilde{r}_{t-1}) = P \cdot \Pr(S_{t-1} = 1 | \tilde{r}_{t-1}) + (1 - Q)[1 - \Pr(S_{t-1} = 1 | \tilde{r}_{t-1})].$$

Next, note that by Bayes' Rule $\Pr(S_{t-1} = 1 | \tilde{r}_{t-1})$ can be written as a function of $\Pr(S_{t-1} = 1 | \tilde{r}_{t-2})$:

$$\Pr(S_{t-1} = 1 | \tilde{r}_{t-1}) = \Pr(S_{t-1} = 1 | \Delta r_{t-1}, \tilde{r}_{t-2})$$

$$= \frac{f(\Delta r_{t-1} | S_{t-1} = 1, \tilde{r}_{t-2}) \Pr(S_{t-1} = 1 | \tilde{r}_{t-2})}{\sum\limits_{i=1}^{2} f(\Delta r_{t-1} | S_{t-1} = i, \tilde{r}_{t-2}) \Pr(S_{t-1} = i, \tilde{r}_{t-2})},$$

where

$$f(\Delta r_{t-1} | S_{t-1} = i, \tilde{r}_{t-2}) = f(\Delta r_{t-1} | S_{t-1} = i)$$

$$= \frac{1}{\sqrt{2\pi h_{it-1}}} \exp \left\{ \frac{-(\Delta r_t - \mu_{it-1})^2}{2 h_{it-1}} \right\}, \qquad i = 1, 2.$$

Hence $p_{1t} = \Pr(S_t = 1 | \tilde{r}_{t-1})$ can be written as a relatively simple nonlinear recursive scheme:

$$p_{1t} = (1 - Q) \left[\frac{g_{2t-1}(1 - p_{1t-1})}{g_{1t-1} p_{1t-1} + g_{2t-1}(1 - p_{1t-1})} \right]$$

$$+ P \left[\frac{g_{1t-1} p_{1t-1}}{g_{1t-1} p_{1t-1} + g_{2t-1}(1 - p_{1t-1})} \right], \tag{A.6}$$

where

$$p_{1t} = \Pr(S_t = 1 | \tilde{r}_{t-1}),$$

$$g_{1t} = f(\Delta r_t | S_t = 1),$$

$$g_{2t} = f(\Delta r_t | S_t = 2).$$

In the case of time-varying Markov switching probabilities, the recursive structure of p_{1t} is the same as for the basic Hamilton model because the regimes follow a first-order Markov process. In this case, however, P and Q are time-varying, depending on some exogenous or predetermined variables, $P_t = g_P(\tilde{W}_{t-1})$ and $Q_t = g_Q(\tilde{W}_{t-1})$, where $g_P(\cdot)$ and $g_Q(\cdot)$ are arbitrary functional forms such that $P_t, Q_t \in (0, 1)$. Diebold, Lee, and Weinbach (1994), for example, use a logistic function. Filardo (1993, 1994) uses these kinds of models in a macroeconomic setting. Recall that in the GRS model $P_t = \Phi(c_1 + d_1 r_{t-1})$ and $Q_t = \Phi(c_2 + d_2 r_{t-1})$.

Having specified the conditional mean functions (μ_{it}) and conditional variance functions (h_{it}) and the dynamics of the switching between regimes, the

log-likelihood function can be written as

$$
\mathcal{L} = \sum_{t=1}^{T} \log \left[p_{1t} \frac{1}{\sqrt{2\pi h_{1t}}} \exp \left\{ \frac{-(\varDelta r_t - \mu_{1t})^2}{2h_{1t}} \right\} \right.
$$
$$
\left. + (1 - p_{1t}) \frac{1}{\sqrt{2\pi h_{2t}}} \exp \left\{ \frac{-(\varDelta r_t - \mu_{2t})^2}{2h_{2t}} \right\} \right]. \tag{A.7}
$$

References

Ball, C.A. and W.N. Torous, 1994, A stochastic volatility model for short term interest rates and the detection of structural shifts, Working paper (Vanderbilt University, Nashville, TN; University of California, Los Angeles, CA).

Bekaert, G. and S.F. Gray, 1995, Target zones and exchange rates: An empirical investigation, Working paper (Duke University, Durham, NC; Stanford University, Stanford, CA).

Berndt, E.K., B.H. Hall, R.E. Hall, and J.A. Hausman, 1974, Estimation and inference in non-linear structural models, Annals of Economic and Social Measurement 4, 653–665.

Bollerslev, T., 1986, Generalized autoregressive conditional heteroskedasticity, Journal of Econometrics 31, 307–327.

Bollerslev, T., R.Y. Chou, and K.F. Kroner, 1992, ARCH modeling in finance, Journal of Econometrics 52, 5–59.

Cai, J., 1994, A Markov model of unconditional variance in ARCH, Journal of Business and Economic Statistics 12, 309–316.

Campbell, J.Y., 1993, Intertemporal asset pricing without consumption data, American Economic Review 83, 487–512.

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 XLVII, 1209–1227.

Cox, J.C., J.E. Ingersoll, and S.A. Ross, 1985, A theory of the term structure of interest rates, Econometrica 53, 385–407.

Diebold, F.X., J-H. Lee, and G.C. Weinbach, 1994, Regime switching with time-varying transition probabilities, in: C. Hargreaves, ed., Nonstationary time series analysis and cointegration (Oxford University Press, Oxford) 283–302.

Engle, R.F., 1982, Autoregressive conditional heteroscedasticity with estimates of U.K. inflation, Econometrica 50, 987–1008.

Engle, R.F., D.M. Lilien, and R.P. Robbins, 1987, Estimating time varying risk premia in the term structure: The ARCH-M model, Econometrica 55, 391–407.

Engle, R.F., V. Ng, and M. Rothschild, 1990, Asset pricing with a factor ARCH covariance structure: Empirical estimates for Treasury bills, Journal of Econometrics 45, 213–238.

Ferson, W.E. and C.R. Harvey, 1991, The variation of economic risk premiums, Journal of Political Economy 99, 385–415.

Filardo, A.J., 1993, The evolution of U.S. business cycle phases, Working paper (Federal Reserve Bank of Kansas City, Kansas City, KS).

Filardo, A.J., 1994, Business cycle phases and their transitional dynamics, Journal of Business and Economic Statistics 12, 299–308.

Friedman, B.M. and D.I. Laibson, 1989, Economic implications of extraordinary movements in stock prices, Brookings Papers on Economic Activity 2, 137–189.

Ghysels, E., 1993, On periodic stochastic regime switching in time series models, Journal of Business and Economic Statistics 12, 289–298.

Gill, P.E., W. Murray, and M.H. Wright, 1981, Practical optimization (Academic Press, Kent).

Gray, S.F., 1995, An analysis of conditional regime-switching models, Working paper (Fuqua School of Business, Duke University, Durham, NC).

Hamilton, J.D., 1988, Rational expectations econometric analysis of changes in regime: An investigation of the term structure of interest rates, Journal of Economic Dynamics and Control 12, 385–423.

Hamilton, J.D., 1989, A new approach to the economic analysis of nonstationary time series and the business cycle, Econometrica 57, 357–384.

Hamilton, J.D., 1990, Analysis of time series subject to changes in regime, Journal of Econometrics 45, 39–70.

Hamilton, J.D., 1994, Time series analysis (Princeton University Press, Princeton, NJ).

Hamilton, J.D. and R. Susmel, 1994, Autoregressive conditional heteroscedasticity and changes in regime, Journal of Econometrics 64, 307–333.

Hansen, B.E., 1992, The likelihood ratio test under nonstandard conditions: Testing the Markov switching model of GNP, Journal of Applied Econometrics 7, S61–S82.

Hentschel, L.E., 1995, All in the family: Nesting symmetric and asymmetric GARCH models, Journal of Financial Economics 39, 71–104.

Hong, C.-H., 1988, The integrated generalized autoregressive conditional heteroscedastic model: The process, estimation and Monte Carlo experiments, Working paper (University of California, San Diego, CA).

Kees, G.K., F.G.J.A. Nissen, P.C. Schotman, and C.C.P. Wolff, 1994, The dynamics of short-term interest rate volatility reconsidered, Working paper (Limburg Institute of Financial Economics, Limburg).

Kim, C.-J., 1994, Dynamic linear models with Markov-switching, Journal of Econometrics 60, 1–22.

Lamoureux, C.G. and W.D. Lastrapes, 1990, Persistence in variance, structural change, and the GARCH model, Journal of Business and Economic Statistics 8, 225–234.

Nelson, D.B., 1990, Stationarity and persistence in the GARCH (1, 1) model, Econometric Theory 6, 318–344.

Nelson, D.B., 1991, Conditional heteroskasticity in asset returns: A new approach, Econometrica 59, 347–370.

White, H., 1980, A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, Econometrica 48, 817–838.