Tree-structured Multiple Regimes in Interest Rates

Francesco Audrino*
University of Lugano

Revised: October 2004

Abstract

This paper develops a generalized tree-structured (GTS) model of the short-term interest rate which accommodates regime-dependent mean reversion and regime-dependent volatility clustering and level effects in the conditional variance. The model is constructed using the idea of multivariate tree-structured thresholds, and nests the popular GARCH and square root processes as simple special cases. It allows us to estimate the optimal number of regimes endogenously from the data and to exploit possible additional information in the term structure and in other macroeconomic variables. We provide empirical evidence of the strong potential of the GTS model in forecasting conditional first and second moments, also in comparison with alternative models of the short rate.

Keywords: Short-term interest rate models; Tree-structured threshold models; Interest rate and macroeconomic variables; Maximum-likelihood estimation.

1 Introduction

Modelling accurately the time-varying dynamics of the short-term interest rate is a crucial issue in financial economics because of its importance in the valuation of interest-rate-sensitive securities and in interest rate risk management. If the final objective is out-of-sample forecasting, studies have to be particularly careful that results are tenable when fitting the different models to real data. This is not the case for many popular models of the short rate, which suffer from

*Address: Institute of Finance, Via Buffi 13, Centrocivico, CH-6904 Lugano, Switzerland. E-mail: francesco.audrino@lu.unisi.ch. Financial support by the National Centre of Competence in Research "Financial Valuation and Risk Management" (NCCR FINRISK) is gratefully acknowledged. The author thanks F. Trojani, P. Porchia and two anonymous referees for helpful comments.

stationarity problems (for example diffusion models) or often imply an explosive conditional variance (for example estimates from GARCH-type models).

Such problems may be due to the fact that the stochastic behavior of the short-term interest rate varies over time. For example, the dynamics of the short rate process during the Federal reserve (Fed) experiment in the 1979-82 period or during the 1973-75 OPEC oil crises in the U.S. seem to suggest some structural breaks in the time series. Consequently, models which involve the estimation of a set of parameters assumed to be fixed over the whole sample period yield misleading and inaccurate results.

For this reason, in the last few years many studies have used different generalizations of the regime-switching model introduced by Hamilton (1989) to describe the short rate process; see, among others, Hamilton (1988), Lewis (1991), Sola and Driffill (1994), Evans and Lewis (1995) and Garcia and Perron (1996). Gray (1996) developed a generalized regime-switching (GRS) model which allows the short rate to exhibit regime-dependent mean reversion and regime-dependent GARCH and level effects in volatility. More recently, Hansen and Poulsen (2000) have extended the short rate model of Vasicek (1977) to include jumps in the local mean.

However, there are at least two additional problems associated with the use of regime-switching models. The first one is that only regime-switching models with at most two or three regime specifications can be used in practice without running into serious computational difficulties. In such models, the optimal number of regimes must be exogenously specified rather than endogenously derived from the data. The second problem that one has to face when using regime-switching models is related to the regime-classification of the observed data. Often the data do not allow clear regime-classifications, that is, the probability of having observed a regime ex-post may hover around a half.

Following the direction taken by regime-switching models, we propose a model for the short-term interest rate process which allows for different regime-dependent conditional mean and variance dynamics. However, the construction of our model is basically different from the regime-switching framework. The model belongs to the threshold autoregressive (TAR) class of models first introduced by Tong (1978), (1990) and Tong and Lim (1980), and generalized to incorporate ARCH effects by Rabenmanjara and Zakoian (1993). In contrast to regime-switching models, in our approach regimes are constructed using the idea of multiple tree-structured thresholds to partition a multivariate predictor space. As we will see, using such a methodology we are able to specify and characterize better the nature of the different regimes, also in connection with some relevant macroeconomic variables for inflation and real activity. In addition, our model allows us, when possible, to interpret the different regimes as periods of contractions or expansions in relation to the business cycles determined by the National Bureau of Economic Research (NBER).

More particularly, to describe the behavior of the short rate process we use a generalization of the tree-structured AR-GARCH model introduced by Audrino and Bühlmann (2001). Thresholds (or splits) are estimated using a binary tree type construction based on the likelihood, where every terminal node represents a (local) AR-GARCH-type process. Our approach allows for a regime-dependent mean reversion and includes regime-dependent heteroskedasticity and level effects in the conditional variance.

Using our approach, we are able to overcome problems arising in the regime-switching framework. Following the procedure introduced by Audrino and Bühlmann (2001) the estimation of the model is computationally feasible also for more than two or three regime specifications. Moreover, by contrast to regime-switching models the optimal number of regimes is derived endogenously and our enables a perfect regime-classification of the observed data.

The inclusion of some exogenous (macroeconomic) variables relevant for prediction in the estimation procedure of our generalized tree-structured model is straightforward. Recent studies have shown that incorporating macroeconomic variables as predictors can improve crucially the estimation of the term structure dynamics; see, among others, Estrella and Mishkin (1997); Evans and Marshall (1998); Ang and Bekaert (2002); Ang and Piazzesi (2003) and Dewachter and Lyrio (2003). For example, Ang and Bekaert (2002) found that regime-switching models incorporating international short rate and term spread information forecast better than univariate regime-switching models. In our model, this additional information can be used to estimate the multivariate thresholds determining the multiple regimes.

Apart from a number of methodological contributions, our study offers important in-sample and out-of-sample empirical results for the description and forecasting of the short rate process. First, using U.S. monthly data we provide strong empirical evidence of the better descriptive and predictive potential of our generalized tree-structured model in comparison to various alternative competitors. In particular, through three different out-of-sample experiments we find that our model significantly outperforms regime-switching models and the realized volatility model introduced by Andersen et al. (1999) and (2001) with respect to various goodness-of-fit statistics for predicting conditional first and second moments. Second, our generalized tree-structured model fitted to the data exhibits in most cases more than two regimes, the optimal number depending on the conditioning information used in the estimation. Similarly to the results found in previous empirical studies, we find two main regimes with the following standard characteristics. The first regime is characterized by low and stable (low volatility) interest rates, with a persistent effect of individual shocks on the conditional variance. The conditional variance is significantly related to the level of the short rate. In the second main regime, interest rates are higher and significantly mean-reverting with respect to a relatively high mean. The effect of individual shocks on the conditional variance is stronger but dies out more quickly. In addition, we collect empirical evidence of the presence of other regimes characterized by a relatively high real activity and a relatively low inflation. Moreover, we find that such regimes can be associated well to different periods of economic expansion and recession as determined by the National Bureau of Economic Research. Third, we find that our model describes quite well the different dynamics of the short rate process during particular event like the 1979-182 period (corresponding to the Fed experiment) , the 1973-75 period (corresponding to the OPEC oil crises) or the years following the crash of October 1987. Fourth, the incorporation in our model of more conditioning information associated with other term structure or macroeconomic variables yields dramatically improved results. In particular, we find that the most relevant predictors are the long-term interest rate (confirming previous results showing that the term structure includes additional information) and two well-known macroeconomic indices: the index known as Help Wanted Advertising in Newspapers (HELP), traditionally thought of as the leading indicator of real activity, and the CPI inflation index.

The rest of the paper is organized as follows. Section 2 presents our generalized tree-structured model and the corresponding estimation procedure. The empirical in-sample results for the time series of monthly U.S. short-term interest rate data are summarized in Section 3. The out-of-sample performance of the models is tested in Section 4 through three different experiments. In the empirical sections (3 and 4), results from our tree-structured models are compared to those from other alternative approaches. Section 5 includes a summary of the main results of the paper and presents our conclusions.

2 Model construction and estimation

This section outlines the generalized tree-structured (GTS) model and describes the model selection procedure that can be applied to it.

2.1 Starting point

For the sake of clarity, it is useful to start by considering a general (nonparametric) model for the short-term interest rate r_t of the form

$$\Delta r_t = \mu_t + \varepsilon_t, \tag{1}$$

where

$$\varepsilon_t = \sqrt{h_t} z_t, \quad \mu_t = g(\Phi_{t-1}), \quad h_t = f(\Phi_{t-1}), \tag{2}$$

for some functions $g(\cdot) \in \mathbb{R}$ and $f(\cdot) \in \mathbb{R}^+$. $(z_t)_{t\in\mathbb{Z}}$ is a sequence of independent identically distributed innovations with zero mean and unit variance. In model (2), the relevant conditioning information, denoted by Φ_{t-1}, is assumed to be as wide as possible. Specifically, we set $\Phi_{t-1} = \{\tilde{r}_{t-1}, h_{t-1}, \mathbf{x}^{\text{ex}}_{t-1}\}$, where $\tilde{r}_{t-1} = \{r_{t-1}, r_{t-2}, \ldots\}$ and $\mathbf{x}^{\text{ex}}_{t-1}$ is a vector of all other relevant exogenous variables used for prediction. In this study, typical factors included in $\mathbf{x}^{\text{ex}}_{t-1}$ are the long-term rate, the spread and some relevant macroeconomic variables such as indices for real activity and inflation. Clearly, such a definition of Φ_{t-1} allows us to exploit all the additional predictive information included in the term structure and in the macroeconomic variables for estimating the dynamics of the short rate process. In particular, the dependence of μ_t on h_{t-1} allows for a (possibly nonlinear) conditional mean effect of volatility. Similarly, the dependence of h_t on $\tilde{r}_{t-1}, h_{t-1}$ and $\mathbf{x}^{\text{ex}}_{t-1}$ allows for a broad variety of asymmetric volatility patterns in reaction to past market and macroeconomic information.

The general model (2) examined in this paper nests several classical models introduced in the literature for the dynamics of the short-term interest rate process. For instance, one immediately sees that Bollerslev's GARCH(1,1) model and a discretized diffusion model motivated by the Cox, Ingersoll and Ross (1985) model are encompassed by (2). In both models, the conditional mean function is parameterized by

$$\mu_t = g(\Phi_{t-1}) = \alpha + \beta r_{t-1}, \tag{3}$$

where α and β are unknown parameters to be estimated. In the discretized diffusion model, the conditional variance is parameterized by

$$h_t = f(\Phi_{t-1}) = \sigma^2 r_{t-1}^{2\gamma}, \tag{4}$$

where σ^2 and γ are the unknown parameters. Similarly, in the GARCH(1,1) model the conditional variance is parameterized by

$$h_t = f(\Phi_{t-1}) = w + a\varepsilon_{t-1}^2 + bh_{t-1}, \tag{5}$$

where w, a and b are the unknown parameters. In the case of the discretized diffusion model $\Phi_{t-1} = \{r_{t-1}\}$, i.e. the only relevant conditioning information is the last lagged level of interest rates. In contrast, the recursive definition of the GARCH(1,1) model implies that the conditional variance depends on the entire history of the data and $\Phi_{t-1} = \{\tilde{r}_{t-1}\}$.

A third classical model introduced in the literature to analyze the short-term interest rate dynamics is the generalized regime-switching (GRS) model proposed by Gray (1996). In Gray's two-regime GRS model, assuming conditional normality within each regime, the conditional mean function is given by

$$\mu_t = g(S_t, \Phi_{t-1}) = p_{t,1}\mu_{t,1} + (1 - p_{t,1})\mu_{t,2} \tag{6}$$

and the variance of changes in the short rate by

$$h_t = f(S_t, \Phi_{t-1}) = p_{t,1}(\mu_{t,1}^2 + h_{t,1}) + (1 - p_{t,1})(\mu_{t,2}^2 + h_{t,2}) - [p_{t,1}\mu_{t,1} + (1 - p_{t,1})\mu_{t,2}]^2, \tag{7}$$

where $p_{t,1}$ denotes the conditional probability to be in regime 1 at time t given the past history $\tilde{r}_{t-1}$, i.e. $p_{t,1} = P[S_t = 1 \mid \Phi_{t-1}]$, and S_t is the unobserved regime at time t. $\Phi_{t-1} = \{\tilde{r}_{t-1}\}$ does not contain S_t or lagged values of S_t. In (6) and (7) the regime-dependent conditional mean functions are parameterized by

$$\mu_{t,j} = \alpha_j + \beta_j r_{t-1},\ j = 1, 2, \tag{8}$$

and the regime-dependent conditional variances by

$$h_{t,j} = w_j + a_j\varepsilon_{t-1}^2 + b_j h_{t-1} + \sigma_j^2 r_{t-1},\ j = 1, 2, \tag{9}$$

where $\alpha_j, \beta_j, w_j, a_j, b_j, \sigma_j^2, j = 1, 2,$ are the unknown parameters. Note that Gray's model, while being of the general form (1), is not encompassed in (2), since conditional means and variances are functions also of unobservable ex-ante probabilities (and not only of an observable information set ϕ_{t-1}).

Our goal is to propose a parametric model for (2) that allows for flexibility in the conditional mean and variance functions g and f while being computationally manageable when applied to real-data examples. As in Audrino and Bühlmann (2001), the basic idea is in the spirit of a sieve approximation of g and f by means of piecewise linear functions. This can be accomplished as follows. We partition the domains of g and f in a finite sequence of regimes (or cells) using a binary tree construction. For any given regime we specify a regime-dependent AR-GARCH type structure for conditional means and volatilities.

2.2 The model

Analogously to the GARCH(1,1) model and the discretized diffusion model introduced in section 2.1, the generalized tree-structured (GTS) model parameterizes the conditional mean $\mu_t(\theta) = g_\theta(\Phi_{t-1})$ and conditional variance $h_t(\theta) = f_\theta(\Phi_{t-1})$ by means of some parametric threshold functions and an unknown parameter vector θ. Our approach follows closely Audrino and Trojani (2003) by incorporating in the threshold definition the additional information deriving from further exogenous (macroeconomic) variables.

The parametric version of model (2) becomes

$$\Delta r_t = \mu_t(\theta) + \sqrt{h_t(\theta)} z_t = g_\theta(\Phi_{t-1}) + \sqrt{f_\theta(\Phi_{t-1})} z_t, \tag{10}$$

for some given parametric functional forms g_θ and f_θ. As in (8) and (9), we model $h_t(\theta)$ by means of a threshold GARCH function f_θ also incorporating level effects as in the square-root process of Cox et al. (1985) and $\mu_t(\theta)$ by means of a threshold, regime-dependent mean-reverting function g_θ. We incorporate in the threshold definitions behind f_θ and g_θ the joint impact of r_{t-1}, ε_{t-1}, h_{t-1} and all other relevant exogenous variables included in Φ_{t-1}.

The construction of g_θ and f_θ follows the structure of a binary tree structured model. It involves a partition $\mathcal{P}$ of the state space G of $\Phi_{t-1} = \{r_{t-1}, \varepsilon_{t-1}, h_{t-1}, \mathbf{x}^{\text{ex}}_{t-1}\}$:

$$\mathcal{P} = \{\mathcal{R}_1, \ldots \mathcal{R}_k\},\ G = \cup_{j=1}^{k} \mathcal{R}_j,\ \mathcal{R}_i \cap \mathcal{R}_j = \emptyset\ (i \neq j)\ .$$

Given a partition cell $\mathcal{R}_j$, we describe the dynamics of r_t on this cell by a local AR(1)-GARCH(1,1) model. This leads to functions for $\mu_t(\theta)$ and $h_t(\theta)$ that depend on (i) the set of parameters of any local AR(1)-GARCH(1,1) model in the generalized tree-structured GARCH model and (ii) the structure of the partition $\mathcal{P}$. More precisely, we have:

$$g_\theta(r,\varepsilon,h,x^{\text{ex}}) = g_\theta^{\mathcal{P}}(r,\varepsilon,h,x^{\text{ex}}) = \sum_{j=1}^{k} (\alpha_j + \beta_j r)\, I_{[(r,\varepsilon,h,x^{\text{ex}})\in\mathcal{R}_j]}, \tag{11}$$

$$f_\theta(r,\varepsilon,h,x^{\text{ex}}) = f_\theta^{\mathcal{P}}(r,\varepsilon,h,x^{\text{ex}}) = \sum_{j=1}^{k} \left(w_j + a_j\varepsilon^2 + b_j h + \sigma_j^2 r\right) I_{[(r,\varepsilon,h,x^{\text{ex}})\in\mathcal{R}_j]}, \tag{12}$$

where $\theta = (\alpha_j, \beta_j, w_j, a_j, b_j, \sigma_j^2;\ j = 1,\dots,k)$.[1] Clearly, $k = 1$ implies a standard AR(1)-GARCH(1,1)-type model which nests the discretized diffusion model specified in (3) and (4) and Bollerslev's GARCH(1,1) model specified in (3) and (5) as special cases. For $k \geq 2$ we obtain a rich class of threshold models, where k also indicate the number of model's regimes.[2] For $k = 2$ or $k = 3$ the GTS model also nests the threshold models introduced by Pai and Pedersen (1999). However, their work was based on a Bayesian approach and they used only information from past short rates for prediction.

Note that the GTS model (10)-(12) is different from the GRS model (6)-(9) proposed by Gray (1996). First of all, in our approach regimes are determined by multivariate tree-structured thresholds. Second, in contrast to regime-switching models the optimal number of regimes is estimated endogenously during the procedure and not given a-priori. Third, the GTS model, being of a threshold type, allows for a perfect regime-classification of the observed data.

The partition $\mathcal{P}$ is constructed on a binary tree where every terminal node represents a rectangular partition cell $\mathcal{R}_j$ whose edges are determined by thresholds. Figure 1 illustrates an

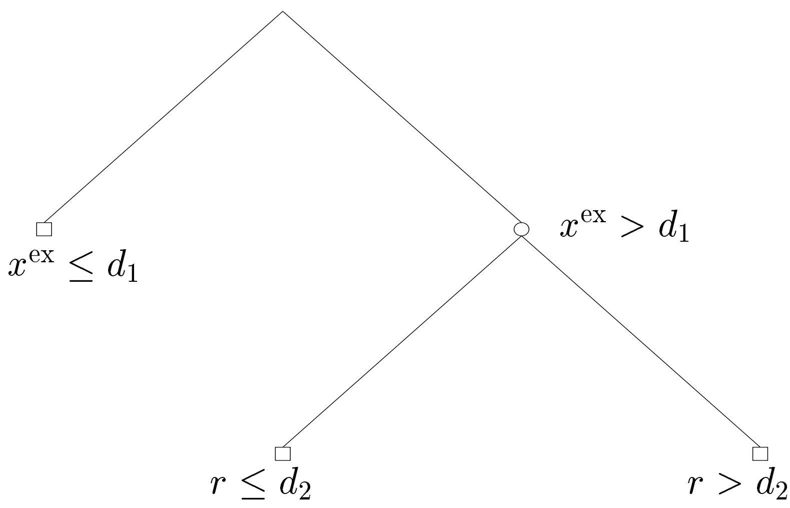

Figure 1: Example of a binary tree partition $\mathcal{P} = \{\mathcal{R}_1, \mathcal{R}_2, \mathcal{R}_3\}$ of the state space $G = \{(r,\varepsilon,h,x^{\text{ex}}); (r,\varepsilon,x^{\text{ex}}) \in \mathbb{R}^3, h \in \mathbb{R}^+\}$.

example of a binary tree partition of the state space

$$G = \{(r, \varepsilon, h, x^{\text{ex}}); (r, \varepsilon, x^{\text{ex}}) \in \mathbb{R}^3, h \in \mathbb{R}^+\} \ ,$$

which involves three partition cells $\mathcal{R}_1$, $\mathcal{R}_2$ and $\mathcal{R}_3$. For simplicity and illustration purposes we assume a one-dimensional exogenous variable x^{ex} identified with the long-term interest rate. Each rectangular partition cell $\mathcal{R}_1, \mathcal{R}_2, \mathcal{R}_3$ corresponds to a terminal node in the tree and determines a regime. The first cell $\mathcal{R}_1 = \{(r, \varepsilon, h, x^{\text{ex}}); x^{\text{ex}} \leq d_1\}$ represents a first regime of r_t in response to low long-term interest rates. The second cell $\mathcal{R}_2 = \{(r, \varepsilon, h, x^{\text{ex}}); x^{\text{ex}} > d_1 \text{ and } r \leq d_2\}$ corresponds to a second regime in response to high long-term interest rates and low short-term interest rates (i.e. high spreads). Finally, $\mathcal{R}_3 = \{(r, \varepsilon, h, x^{\text{ex}}); x^{\text{ex}} > d_1 \text{ and } r > d_2\}$ represents a third regime in response to both high short- and long-term interest rates.

It is important to bear in mind that the threshold values d_1, d_2 are not restricted and are jointly estimated. Thus, low (high) interest rates with respect to the partition $\{\mathcal{R}_1, \mathcal{R}_2, \mathcal{R}_3\}$ is not stated in absolute terms but rather to mean interest rates that are sufficiently below (above) the threshold values d_1 and d_2.

The negative log-likelihood[3] for model (10) is

$$-\ell(\theta; \Phi_2^n) = -\sum_{t=2}^{n} \log \left[\sqrt{h_t(\theta)}^{-1} p_Z \left((\Delta r_t - \mu_t(\theta)) / \sqrt{h_t(\theta)} \right) \right], \tag{13}$$

where $p_Z(\cdot)$ is the density function of the distribution of the standardized innovation z_t and $\Phi_2^n = \{\Phi_2, .., \Phi_n\}$. Therefore, for any given partition $\mathcal{P}$ model (10) can be estimated by means of (pseudo) maximum likelihood. The choice between different partition structures (i.e. the selection of the optimal threshold functions) involves a model choice procedure for non-nested hypotheses.

Summarizing, we may say that a flexible procedure for the estimation of the GTS model is as follows.

(i) For any given partition $\mathcal{P}$ the estimation of θ is performed by (pseudo) maximum likelihood based on a gaussian (pseudo) log likelihood[4] and the parametric forms (11) and (12) for g_θ and f_θ.

(ii) Model selection of the optimal threshold function (i.e. the optimal partition $\mathcal{P}$) is performed via a tree-structured partial search[5]. Within any data-determined tree structure the optimal model is finally selected according to the Bayesian-Schwarz Information Criterion (BIC).

We present some more details on these two issues in the next section.

2.3 The estimation procedure

The estimation of the GTS model (11)-(12) by means of the general steps (i), (ii) is achieved as follows. We estimate a largest generalized tree-structured model, given a maximal number of candidate thresholds. Then, we apply a model selection procedure for non-nested models that selects an optimal subtree of the largest tree estimated in the first step. This estimation procedure follows closely the one introduced by Audrino and Bühlmann (2001).

2.3.1 Growing the "maximal" binary tree

We first fix a maximal allowed number $M + 1$ of partition cells in the tree. This corresponds to the maximal number of possible multivariate regimes in conditional means and variances. For any coordinate axis of the multivariate prediction space G that has to be split we search for multivariate thresholds over grid points that are empirical α-quantiles of the data along the relevant coordinate axis. Typically, we fix the empirical quantiles as $\alpha = i/\text{mesh}$, $i = 1, .., \text{mesh} - 1$, where mesh determines the fineness of the grid on which we search for multivariate thresholds. Typically, we choose mesh = 8.

The partition of the prediction space G of $\Phi_{t-1} = \{r_{t-1}, \varepsilon_{t-1}, h_{t-1}, \mathbf{x}^{\text{ex}}_{t-1}\}$ into a maximal number of $M + 1$ cells is then performed as follows. A first threshold $d_1 \in \mathbb{R}$ or $\mathbb{R}^+$ in one component indexed by a component index $\iota_1 \in \{1, \ldots, p\}$, with p the dimension of G, partitions G as

$$G = \mathcal{R}_{left} \cup \mathcal{R}_{right},$$

where $\mathcal{R}_{left} = \{(r, \varepsilon, h, \mathbf{x}^{\text{ex}}) \in \mathbb{R}^2 \times \mathbb{R}^+ \times \mathbb{R}^{p-3};\ (r, \varepsilon, h, \mathbf{x}^{\text{ex}})_{\iota_1} \leq d_1\}$ and $(r, \varepsilon, h, \mathbf{x}^{\text{ex}})_{\iota_1}$ is the ι_1 component of the vector $(r, \varepsilon, h, \mathbf{x}^{\text{ex}})$. $\mathcal{R}_{right}$ is defined analogously using relation '$>$' instead of $\leq$. In a second step, one of the partition cells $\mathcal{R}_{left}$, $\mathcal{R}_{right}$ is again partitioned with a second threshold d_2 and a second component index ι_2 in the same way as above.

We iterate this procedure. Specifically, for the $m-$th iteration step, we specify a new pair (d_m, ι_m) (which determines a new threshold for the component indexed by ι_m) and an existing partition cell which is going to be further split into two subcells. For a new pair $(d, \iota) \in \mathbb{R} \times \{1, \ldots, p\}$ refinement of an existing partition $\mathcal{P}^{(old)}$ is obtained by picking $\mathcal{R}_{j^*} \in \mathcal{P}^{(old)}$ and splitting it as

$$\mathcal{R}_{j^*} = \mathcal{R}_{j^*,left} \cup \mathcal{R}_{j^*,right} \ . \tag{14}$$

This gives a new (finer) partition of G as

$$\mathcal{P}^{(new)} = \{\mathcal{R}_j, \mathcal{R}_{j^*,left}, \mathcal{R}_{j^*,right}, j \neq j^*\}, \tag{15}$$

where (d, ι) describes a threshold and a component index such that

$$\mathcal{R}_{j^*,left} = \{(r, \varepsilon, h, \mathbf{x}^{\text{ex}}) \in \mathcal{R}_{j^*} \subset G; (r, \varepsilon, h, \mathbf{x}^{\text{ex}})_\iota \leq d\}. \tag{16}$$

$\mathcal{R}_{j^*,right}$ is defined analogously by the '$>$' instead of '$\leq$' relation. The whole procedure finally determines a partition $\mathcal{P} = \{\mathcal{R}_1, \ldots, \mathcal{R}_k\}$ which can be summarized by a binary tree, where the terminal nodes represent the rectangular partition cells in $\mathcal{P}$, see again Figure 1.

To select the specific threshold and component index (d, ι) in each iteration step of the above procedure we proceed by optimizing the conditional negative (pseudo) log-likelihood (13). Details on the implied algorithm are given in Audrino and Bühlmann (2001). At the end we obtain a partition $\mathcal{P}^{(M)}_{opt}$ corresponding to a large binary tree equipped with parameter estimates $\hat{\theta}^{(M)}$.

To end this section it is important to draw attention to the fact that only the number of estimated thresholds (or, equivalently, the number of estimated regimes) in the maximal tree can be arbitrary (high). The maximal partition cannot be. Since similarly to the classification and regression trees procedure (CART, see Breiman et al., 1984) for computational feasibility it is not possible to estimate the best partition among all the possible ones, the proposed procedure is based on a hierarchical strategy. The first split in the binary tree is the one that generates the greatest improvement in minimizing the conditional negative (pseudo) log-likelihood (13) (i.e. the statistical criterium chosen for predictive accuracy). Then, the second split in the tree is the one that added to the (already fixed) first one yields the maximal improvement in minimizing the conditional negative log-likelihood, and so on. At the end, we get a "unique" data driven maximal partition that is optimal with respect to the conditional negative log-likelihood criterium. In other words, the procedure yields a natural ordering of the partition cells (regimes) based on the optimal binary tree structure.

2.3.2 Selection of the optimal number of regimes (subtree) by pruning

The maximal binary tree (or equivalently the maximal partition $\mathcal{P}^{(M)}_{opt}$) constructed with the above estimation procedure can be too large (or too fine, respectively). This can lead to overfitting and poor out-of-sample goodness-of-fit results. As with CART, we correct by pruning. Specifically, we search for a best subtree of $\mathcal{P}_{opt}^{(M)}$ and consequently we estimate the optimal number of regimes using the BIC model selection criterion (17) below. Note that the selection of the optimal subtree, too, is entirely data driven.

Let τ be the set of all binary subtrees of $\mathcal{P}_{opt}^{(M)}$ and denote for brevity an arbitrary element of τ by $\mathcal{P}_i$. Thus, every subtree $\mathcal{P}_i$ corresponds to a partition of G which can be represented as a binary tree. For every $\mathcal{P}_i$ we compute the implied pseudo maximum likelihood estimate $\hat{\theta}^{\mathcal{P}_i}$, according to (13) and based on functions $g_{\hat{\theta}^{\mathcal{P}_i}}^{\mathcal{P}_i}$, $f_{\hat{\theta}^{\mathcal{P}_i}}^{\mathcal{P}_i}$ of the form (11) and (12). We then consider the penalized negative log-likelihood (or BIC) statistic

$$\mathrm{BIC}(\mathcal{P}_i) = -2 \cdot \ell\big(\hat{\theta}^{\mathcal{P}_i}; \Phi_2^n\big) + \dim(\hat{\theta}^{\mathcal{P}_i}) \cdot \log(n-1) \tag{17}$$

as a measure of predictive performance for the given partition $\mathcal{P}_i$. We finally select the binary tree (or equivalently the partition) $\hat{\mathcal{P}}$ that minimizes (17) across all subtrees of the maximal tree $\mathcal{P}_{opt}^{(M)}$. The resulting estimated GTS model is given by a parametric functional form (10) with functions $g_{\hat{\theta}^{\hat{\mathcal{P}}}}^{\hat{\mathcal{P}}}$ and $f_{\hat{\theta}^{\hat{\mathcal{P}}}}^{\hat{\mathcal{P}}}$ in (11) and (12) based on the resulting optimal partition $\hat{\mathcal{P}}$.

2.3.3 Consistency and reliability of the estimation procedure

Proofs of consistency of the model selection procedure for the case that the true model is in the class of GTS models are very difficult to obtain. Analogously to CART, it is possible to prove theorems that study the behavior of the prevailing parameter estimators when growing the tree. However, such results do not imply model selection consistency either. Furthermore, it is quite hard to believe that the "correct" generating process in our and similar real data examples is indeed a tree-structured AR-GARCH-type model. For this reason, it is more important to prove consistency of the estimates in a tree-structured AR-GARCH model under a possible model mis-specification, rather than showing consistency of the model selection strategy under the assumption of a correctly specified tree-structured model. Such consistency results can be found in Audrino and Bühlmann (2001) for the volatility estimates.

Moreover, we verify the reliability of the tree-structured partitioning procedure proposed in section (2.3.1) on various simulated data samples. Most of the work was already done by Audrino and Bühlmann (2001). Their simulations (see data samples 1 to 5) show that the estimation procedure can pick up the true partition of the predictor space within a reasonable approximation when the data generating process is of a tree-structured AR-GARCH type.

In this study, we investigate what kinds of partitions of the predictor space are chosen by the tree-structured estimation procedure when the true data generating process is the two-regime regime switching model proposed by Gray (1996). We simulate 50 independent realizations for the level and changes of the short-term interest rate process from the two-regime generalized regime switching model (6)-(9) with parameters given in Table 3. We examine realizations of 500 monthly observations to be consistent with our real data investigation of section (3.1). In our simulations we regard simulated interest rate levels and changes, estimated innovations and conditional variances as candidates for possible threshold variables.

In all 50 simulations we find that the optimal GTS model has two different regimes characterized by low and stable interest rates, and high interest rates with high volatility. As expected, the only significant predictor variable is always the level of the short rate. Optimal threshold values depend on the simulation but are in most cases one of the upper quantiles of the simulated short-term interest rate levels. This is reasonable since it means that most of the time we are in the low and stable short rate regime. In short, the estimated optimal partitions for all 50 simulations are consistent with the true data generating process.

We also investigate the performance of the conditional mean and variance estimates obtained from the GTS model, in comparison with those from single-regime and two-regime generalized regime switching models. To measure the goodness-of-fit of the different models we consider the classical mean absolute error (MAE) and root mean squared error (RMSE) statistics. MAE and RMSE average values across the 50 independent simulations for conditional mean and variance estimates are summarized in Table 1.

TABLE 1 ABOUT HERE.

Table 1 shows quite well the good potential of the GTS model in estimating conditional first and second moments. In fact, the GTS model clearly outperforms the single-regime GRS model with respect to all performance measures. Moreover, the GTS model performs well also in comparison to the true data generating two-regime GRS model.

3 Data and estimation results

3.1 Data

The data used in this study are monthly one-month U.S. Treasury bill rates downloaded from the Fama CRSP Treasury bill files. The data span the time period between January 1960 and

December 2001, for a total of 504 observations. Figure 2 plots the data as well as the monthly changes in short-term interest rates. Table 2 presents some sample statistics.

FIGURE 2 AND TABLE 2 ABOUT HERE.

Figure 2 illustrates quite aptly the dramatic changes in the short-term interest rates that occurred during the OPEC oil crises in the 1973-75 period and the Fed experiment in the 1979-82 period. The volatility of the monthly changes associated with the Fed experiment is striking. Volatility is also noticeably higher than average during the 1973-75 period and immediately after the October 1987 stock market crash. As expected, Table 2 shows that the mean change in the short-term interest rates is close to zero, that there is significant excess kurtosis, and that the correlation between Δr_t and r_{t-1} is negative. All these stylized facts have been documented elsewhere.

To exploit the possible additional information included in the yield curve, we also downloaded the 60-month zero coupon bond rates from the Fama CRSP discount bond files. Some sample statistics for the 60-month yields, as well as for the spread between long- and short-term rates, are summarized in Table 2.

We also use additional macroeconomic variables as conditioning predictors in our generalized tree-structured model, since we believe that they can substantially improve estimation and forecast. We divide the macroeconomic variables into two main groups. The first group consists of two inflation measures based on the CPI and the PPI of finished goods. The second group contains variables that capture real activity: the index of Help Wanted Advertising in Newspapers (HELP), unemployment (UE) and the growth rate of industrial production (IP). In addition, we also consider monthly log-returns of the S&P500 index. All the macroeconomic data have been downloaded from *Datastream International* for the time period under investigation. This list of variables includes most variables that have been used in the macro literature. Among these variables, CPI and HELP are traditionally thought of as leading indicators of inflation and real activity, respectively. Summary statistics of these variables are reported in Table 2.

3.2 Generalized regime-switching (GRS) model estimation results

We begin our analysis by considering a single-regime and a two-regime GRS models, as in Gray (1996), given by (6)-(9), since they can be used as benchmark models to test the accuracy of the generalized tree-structured GARCH model. In particular, the single-regime model is included in the generalized tree-structured GARCH construction as a simple special case. The parameter estimates for these models appear in Table 3. Results are computed for the whole time period beginning January 1960 and ending December 2001, for a total of 504 monthly observations. The detailed specification of the models appears below the table.[6]

TABLE 3 ABOUT HERE.

The first column of Table 3 reports the estimates for the single-regime version of the GRS model. The mean reversion parameter (β_1) is, as expected, negative and significant. The implied long-run mean ($-\alpha_1/\beta_1$) is 3.78% per annum. The GARCH effect appears to be important in characterizing conditional variances. On the contrary, the CIR parameter σ_1^2 is found to be statistically insignificant. The most important factors in determining volatility are recent volatility and shocks. In contrast to previous studies on weekly data, in our case diagnostic tests indicate that the single-regime GRS model does a good job in modelling the stochastic volatility of short-term interest rates. The Ljung-Box statistics relating to the squared standardized residuals indicate no significant serial correlation. Moreover, in our case the assumption of stationarity is not violated: $\hat{a}_1 + \hat{b}_1 = 0.9218 < 1$.

The second column of Table 3 reports estimates of the two-regime GRS model. The estimates of the conditional mean and variance parameters are similar to those found by Gray (1996) and confirm a certain asymmetry across regimes. In the high volatility and interest rates regime (regime 1) there is weak evidence of mean reversion. On the contrary, there is no evidence whatsoever of mean reversion at low and moderate interest rates (coinciding with episodes of regime 2 and lower volatility). As in Gray (1996), the conditional variance appears to separate into a GARCH and a CIR regime. In regime 1 the CIR parameter yield the most significant (although short of 1% confidence level) effect driving the conditional variance. In regime 2, the most important factor in determining volatility is recent volatility, whereas the parameter estimate for the level effect is approximately zero.

Within each regime, the GARCH processes are stationary and less persistent than in the single-regime model ($a_i + b_i < 0.71,\ i = 1, 2$). Conversely, allowing for regime switches has substantially increased the value of the CIR parameter in the high volatility regime. This illustrates a potential advantage of the regime-switching model over the single-regime model. In the regime-switching model, volatility clustering can be caused by three factors, whereas in the single-regime the only source of clustering lies in the GARCH process. Since the single-regime model cannot capture the persistence of regimes, all 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.[7]

As expected, the GRS model does a relatively good job in modelling the stochastic volatility of short-term interest rates. The Ljung-Box statistics relating to the squared standardized residuals indicate no remaining serial correlation.

The top panel of Figure 3 contains a plot of the ex-ante and smoothed probabilities from the two-regime GRS model.

FIGURE 3 ABOUT HERE.

This figure points to at least two periods during which the process was most likely in the high-variance regime: 1979-82 and late 1987-88. These periods may be explained quite intuitively. The first one is clearly driven by the Fed experiment. The second period corresponds to the months immediately after the stock market crash of October 1987. Moreover, we also observe some spikes of short duration during the OPEC oil crisis of 1973-75, a period known for its increased financial market volatility. However, the regime classification cannot be given for sure since the average probability of this period's being in regime 1 is only about 0.2. This is one of the disadvantages of working with regime-switching models: a perfect regime-classification of the data is in some cases not allowed. The bottom panel of Figure 3 contains a plot of the conditional standard deviation implied by the two-regime GRS model. Once again, the periods of high volatility are particularly apparent.

3.3 Generalized tree-structured (GTS) model estimation results

In this section, we analyze the parameter estimates of the conditional mean and variance of the short-term interest rate for two different forms of the generalized tree-structured (GTS) model (10) with conditional mean and variance equations specified by (11) and (12).

3.3.1 Simple GTS model

We begin the analysis of generalized tree-structured (GTS) models by considering a model which uses only endogenous information for prediction (i.e. not exploiting the possible additional information included in the term structure or in other macroeconomic variables). We call such model simple GTS model. The parameter estimates for this model appear in Table 4. Results are computed for the whole time period beginning January 1960 and ending December 2001, for a total of 504 monthly observations. The detailed specification of the model is shown below the table.

TABLE 4 ABOUT HERE.

We find that the simple GTS model has four different regimes, endogenously estimated in our procedure. The first two regimes are characterized by low and stable interest rates. However, the short rate interest rate process behaves differently in these two regimes in response to past negative and positive[8] interest rate changes. In the first regime the tendency of the interest rate process is positive. Negative past interest rate changes act as warning signals that the behavior of the short-term interest rate process may change, in reaction to a different monetary policy or to new real activity conditions. The implied long-run mean is 6.73% and, since it lies outside the regime specification, acts as a kind of high external attractor. In this regime the conditional mean expected change is always positive, although the strength of the attraction of the external high long-run mean is low. In fact, when the short rate is 3%, the conditional mean change is 5.33 basis points. When past interest rate changes are negative, the level of the interest rate tends to increase and the process tends to switch to another regime associated with higher interest rates. In this regime the effect of individual shocks is not large immediately, but significantly persistent.

In the second regime, there is weak evidence of a slow negative mean reversion with respect to a relatively low implied long-run mean (3.19%), which acts as a kind of reflecting barrier since it is rare in our sample to be below 3.19%. However, the strength of this reflection is relatively moderate. When the short rate is 4% the expected change is 4.35 basis points. In this regime, positive past interest rate changes act as good signals from the market. Once the interest rate falls below the implied long-run mean level, the process switches to regime 1 since we have a negative past interest rate change. As with regime 1, the GARCH parameter estimate in the conditional variance is highly significant. What is different here is that there is also weak evidence that the conditional variance is related to the level of the short rate. Note that, although the value of the CIR parameter is small, it is also economically significant.

The third regime is characterized by moderate to high interest rates and volatilities. In this regime there is statistical evidence of negative mean reversion. The implied long-run mean is 5.68% and acts as a reflecting barrier. However, the strength of this reflection is moderate: when the short rate is 5% and 6.5% the conditional mean change is -4.66 and 5.69 basis points. This regime acts like a temporary regime: when the short rate is above (below) the implied long-run mean, it tends to increase (decrease) further and to move to the high (low) interest rate regime 4 (regimes 1 and 2). Both CIR and GARCH parameters are found to be statistically significant.

The fourth and last regime is characterized by high interest rates and volatilities. Analogously to regime 1, the implied long-run mean is 6.91% and, since it lies outside the regime specification, acts as a kind of low external attractor. In this regime the conditional mean expected change is always negative, with a tendency to reduce the level of the short rates. Unlike regime 1, the speed of the conversion to the external low implied long-run mean is high. Moreover, the parameter estimates of the conditional mean are found to be highly statistically significant. Individual shocks have a large immediate and persistent effect on the conditional variance. Within this regime, the GARCH process in not stationary ($a_4 + b_4 > 1$). However, as we will see more clearly in Figure 4, this regime characterizes only particular events like the 1979-82 Fed experiment.

In sum, we find statistical evidence of the existence of more than two asymmetric threshold regimes. The behavior of the short rate process is different in each regime. In contrast to single-regime models, using our approach we are able to separate the different sources of volatility persistence. Moreover, similarly to the two-regime GRS model our results suggest that the mean reversion, which is assumed by many models to be operating continuously, in fact does not occur in the different regimes. However, the global behavior of the conditional mean estimated using the simple GTS model is similar to that of a mean-reverting process.

As expected, the simple GTS model is quite suitable for modelling the stochastic volatility of short-term interest rates. The Ljung-Box statistics relating to the squared standardized residuals indicate no remaining serial correlation. The top panel of Figure 4 contains a plot of the regime classification of the observations from the simple GTS model. The detailed specification of each regime appears below the figure.

FIGURE 4 ABOUT HERE.

Although we have seen that regime 1 and regime 2 have different economic and statistical characteristics, they are both associated with time periods of low and stable interest rates. This is also shown in Figure 4. For example, observations belonging to the 1961-69 and the 1991-2001 time periods are in general classified in regimes 1 and 2 according to the simple GTS model. Considering the business cycle expansions and contractions determined by the Business Cycle Dating Committee of the National Bureau of Economic Research, most time periods coinciding with regimes 1 and 2 are associated with long expansions. The temporary regime 3 characterized by moderate to high interest rates and volatilities is associated with particular episodes like the 1973-75 OPEC oil crises and the years after the market crash of October 1987. During these time periods more than 75% of the observations are classified in regime 3. Finally, the 1979-82 Fed experiment and the short peak corresponding to the end-year of October 1987 are episodes classified in the high-volatility regime 4.

The bottom panel of Figure 4 contains a plot of the conditional standard deviation implied by the simple GTS model superimposed on those from a two-regime GRS fit, for comparison purposes. Once again, periods of high, moderate and low volatility are particularly apparent. Volatility dynamics from both models are similar. Summarizing, Figure 4 highlights that particular episodes and different business cycles coincide with the different regimes estimated using the simple GTS model. In contrast to the two-regime GRS model, we find that observations occurring at the time of the Fed experiment and of the OPEC oil crises are not classified in the same regime.

3.3.2 Full GTS model

To end our analysis, we consider a full generalized tree-structured (GTS) model which also uses the possible additional information included in the term structure and in other relevant macroeconomic variables for prediction. All variables considered in the estimation of the full GTS model are listed in Table 2. The parameter estimates for this model appear in Table 5. Results are computed for the whole time period beginning January 1960 and ending December 2001, for a total of 504 monthly observations. The detailed specification of the model in noted under Table 5.

TABLES 2 AND 5 ABOUT HERE.

Once again we find that the estimated full GTS model has four regimes. However, the regimes are characterized differently from those estimated in the simple GTS model and lead to different economic and statistical interpretations.

The first regime is characterized by a low real activity.[9] Typical episodes in regime 1 are periods of low and stable interest rates. In this regime, the implied long-run mean is relatively low (3.26%) and there is statistical evidence of mean reversion. The speed of the reversion is moderate: when the short rate is 2% and 6% the conditional mean change is 6.12 and -13.24 basis points. As in the case with the low-volatility regime in the two-regime GRS model, individual shocks have a small immediate effect on the conditional variance, but are strongly persistent. Analogously to regime 2 in the simple GTS model, the conditional variance is also significantly related to the level of the short rate. However, in this case, the small value of the CIR parameter renders it economically insignificant.

The second and third regimes are characterized by high real activity and low inflation. Yet, the short-term interest rate process behaves differently in these two regimes in response to past negative and positive interest rate changes. The conditional mean behavior in regimes 2 and 3 is essentially the same. In both regimes, there is statistical evidence of a fast mean reversion around a moderate implied long-run mean (4.66% and 3.74% per annum in regime 2 and 3, respectively). Nevertheless, the conditional variance dynamics are completely different. In regime 2 individual shocks have an immediate impact on the conditional variance, but die out quickly, whereas in regime 3 individual shocks have a small immediate effect on the conditional variance but are significantly persistent. This difference can be economically interpreted as follows. In regime 2, past negative interest rate changes may act as warning signals that the behavior of the short-term interest rate process can change. This may be due to a different monetary policy affecting the level of the short rate and on inflation, and/or to a worsening of the production conditions with a consequent reduction of real activity. In other words, the uncertainty about the future behavior of the interest rate process increases, and so does the immediate impact of individual shocks. In contrast, in regime 3 past positive interest rate changes are seen as good signals. For this reason, the impact of individual shocks is not large immediately, but strongly persistent. In both regimes, the CIR parameter is not statistically and economically significant.

The fourth and last regime is characterized by high real activity and high inflation. In this regime, interest rates and volatility are generally high. There is strong evidence of mean reversion around a relatively high implied long-run mean (6.39%). The speed of the reversion is high, although not so high as in regimes 2 and 3. Similarly to the high-volatility regime 4 in the simple GTS model, the conditional variance is not significantly related to the level of the short rate. Individual shocks have a moderate immediate impact and are strongly persistent, although less than in the low-volatility regime 1.

As expected, the full GTS model is very appropriate for modelling the stochastic volatility of the short-term interest rate process. The Ljung-Box statistics relating to the squared standardized residuals indicate no remaining serial correlation. The top panel of Figure 5 contains a plot of the regime classification of the observations from the full GTS model. The specification of each regime is detailed below the figure.

FIGURE 5 ABOUT HERE.

Figure 5 clearly shows that regime 1 is associated with time periods of low and stable interest rates. Although we have seen that regime 2 and regime 3 have different economic and statistical characteristics, they both appear in the same two time periods characterized by low inflation and high real activity. Observations belonging to the 1985-87 and the 1994-2000 time periods are classified in regimes 2 and 3 according to the full GTS model. Considering once again the business cycle expansions and contractions determined by the Business Cycle Dating Committee of the National Bureau of Economic Research, we find, as expected, that these two time periods coincide with long expansions. Finally, the first half of the 1973-75 period, the 1979-82 Fed experiment and the years after the stock market crash of October 1987 are episodes classified in the high-volatility, high inflation and high real activity regime 4.

The bottom panel of Figure 5 contains a plot of the conditional standard deviation implied by the full GTS model superimposed on those from a two-regime GRS fit for comparison. Once again, periods of high and low volatility are particularly apparent. When confronting the volatility estimates from the full GTS model with those from the two-regime GRS model, we see some important differences in the volatility dynamics, in particular during the 1979-82 Fed experiment period and the months following the stock market crash of October 1987. This is a consequence of the different approaches used to estimate the regimes in the models. As we will see in the following out-of-sample experiments, these differences will prove crucial in issuing accurate volatility forecasts.

4 Three out-of-sample experiments

In this section, we investigate whether the introduction of multiple regimes in the GTS models brings to overfitting. This can be easily determined by performing a series of out-of-sample tests. Moreover, using such type of tests we are also able to establish the economic significance of allowing for more than one regime. In performing the out-of-sample tests, we estimate the parameters of each particular model over an in-sample period and compute time series of conditional mean and variances over a subsequent out-of-sample period, holding the estimated parameters and regime structure fixed.

We always compare goodness-of-fit results of the simple and full GTS models with those of (i) single-regime and two-regime GRS models, (ii) an extended two-regime GRS model where the time-varying transition probabilities are also allowed to depend on further exogenous (macroeconomic) variables[10], and (iii) the realized volatility AR-GARCH model as proposed (among others) by Andersen et al. (1999) and (2001). To estimate the realized volatility AR-GARCH model we download daily one-month interest rate data from the CRSP database. Unfortunately, daily data have been collected only since January 1970. For this reason, our out-of-sample performance tests cover the time period between January 1970 and December 2001, for a total of 7979 daily observations and 384 monthly observations.

By analogous with Gray (1996), we quantify the goodness-of-fit of the different models for estimating and predicting monthly conditional first and second moments by means of various measures. Since the conditional variance is an expectation of squared innovations to the interest rate process, we compare volatility estimates from the different approaches to the actual squared innovations. The difference between volatility estimates and actual squared innovations is computed for the in-sample estimation period and for the out-of-sample backtesting period. This difference is then summarized in the form of root mean squared errors (RMSE), mean absolute errors (MAE) and the R^2 between actual volatility and estimated volatility. Mathematically speaking, we consider

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{h}_t - (\Delta r_t - \hat{\mu}_t)^2\right)^2}, \tag{18}$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|\hat{h}_t - (\Delta r_t - \hat{\mu}_t)^2\right| \quad \text{and} \tag{19}$$

$$R^2 = 1 - \frac{\sum_{t=1}^{n}\left(\hat{h}_t - (\Delta r_t - \hat{\mu}_t)^2\right)^2}{\sum_{t=1}^{n}(\Delta r_t - \hat{\mu}_t)^4}. \tag{20}$$

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

Since we are also interested in the accuracy of the different models in predicting conditional first moments, we also compute classical in-sample and out-of-sample MAE and RMSE statistics for the estimated innovations $\hat{\epsilon}_t = \Delta r_t - \hat{\mu}_t$. In addition to these performance measures, we also consider the negative log-likelihood computed for the out-of-sample period.

TABLES 6 AND 7 ABOUT HERE.

Tables 6 and 7 show performance results of three different out-of-sample tests for conditional mean and volatility predictions constructed using the single, two-regime and extended two-regime GRS models, the simple and full GTS models and the realized volatility AR-GARCH model.

Analogously to Gray (1996), in the first test the models are estimated from the start of the sample up to the Fed experiment, a period that includes the OPEC oil crises. The out-of-sample period includes the whole Fed experiment, the 1987 market crash an the subsequent period characterized on average by low volatility through to the end of the sample. The optimal GTS models have two regimes: the simple GTS model in response to past short-term interest rates above and below a 8.0125 threshold; the full GTS model in response to past HELP values above and below a 86.16 threshold value. Both GTS models perform well in this experiment, with a little advantage of the full GTS model over the simple GTS model in particular for predicting conditional first moments. The parameters of the high-volatility regime in both models are well-identified by the oil shock. In contrast, two-regime GRS models performs extremely badly in this experiment. They predict too high levels of volatility in the low-volatility regime, whereas interest rates turn out to be very stable. This results in high forecast errors and a negative R^2. Moreover, both models are clearly outperformed by the single-regime GRS model over the out-of-sample period. This problem does not seem to depend on overfitting, since the in-sample statistics, too, are better for the single-regime GRS model. In contrast, it can be due to the very short estimating period consisting only of 108 monthly observations. Something similar happens with the realized volatility model: it yields no better out-of-sample results than the single-regime GRS model. Differences between the models are smaller when considering performance measures for conditional first moments. Only the full GTS model shows some advantage over the single-regime GRS model.

For the second test, the models are estimated over the first half of the sample, and predictions are attempted over the second half of the sample. The in-sample estimation period includes the 1973-75 period (OPEC oil crises) and part of the Fed experiment. The out-of-sample period includes the remainder of the Fed experiment and the stock market crash of October 1987. In this experiment, the in-sample period is significantly more volatile than the out-of-sample one (sample variance of the short rate changes in the two periods is 1.037 and 0.242, respectively). The simple GTS model has two optimal regimes in response to past short-term interest rate levels. The full GTS model has 3 optimal regimes in response to past HELP and CPI values and clearly outperforms all the alternative competitors. Similarly to the first test, all the GRS models show overfitting problems and yield bad volatility predictions. Extending the GRS model to incorporate additional (macroeconomic) information achieves some significant out-of-sample improvements over the single-regime and two-regime GRS models. Nevertheless, the extended two-regime GRS specification shows a negative R^2, too. It seems here that two regimes are not enough to forecast accurately the conditional variances of the interest rate process. The realized volatility model works quite well, but is clearly outperformed by the full GTS model. Like before, differences in forecasting conditional first moments are smaller: only the full GTS model shows some significant advantage over all competitors.

The final test examines the short-term forecasting ability of the model in estimating the models over the entire sample except for the last 8 years. The in-sample period is once again significantly more volatile than the out-of-sample period (sample variance of the short rate changes in the two periods is 0.693 and 0.081, respectively). Thus, the sample variance of the short rate changes is more than 8 times larger in the in-sample period than in the out-of-sample period. Once again the optimal full GTS model has more than two regimes (three in this case) in response to past short- and long-term interest rate levels. All the models perform well in this experiment, except for the single-regime GRS model which shows a negative R^2. The extended GRS and the full GTS models show the best out-of-sample results.

In summary, the GTS models perform well in all three tests with respect to four different performance measures for conditional first and second moments. The full GTS model that also uses additional (macroeconomic) information for prediction clearly outperforms all the alternative competitors (including the simple GTS model). In contrast, in some cases the GRS models yield in predictions that are not accurate enough. This shortcoming is not corrected when incorporating in the GRS model additional exogenous information for prediction. Our full GTS model also outperforms the realized volatility AR-GARCH model that uses information of daily recorded one-month short rate data for predicting monthly conditional first and second moments. The realized volatility model works quite well in the different out-of-sample experiments, but, since it assumes fixed parameters for the whole sample, is not able to capture completely the time-varying behavior of the short rate process. We conclude that the description of the short rate process switches in general between more than two regimes characterized by past values of multivariate endogenous and exogenous (macroeconomic) predictor variables lying below (or above) some estimated thresholds. Moreover, volatility forecast from the classical GRS models can be considerably improved using the GTS approach and may be profitably employed in the valuation of interest-rate-sensitive securities and interest rate management.

5 Conclusions

We proposed a generalized tree-structured (GTS) model of the short-term interest rate which allows for an endogenous number of estimated multiple regimes. The short rate exhibits a different degree of mean reversion and a different form of conditional heteroskedasticity in each regime. The regime-dependent conditional variance function is very general and nests the classical GARCH and CIR (square root process) models as simple special cases. In the GTS approach, regimes are constructed using the idea of multivariate tree-structured thresholds, estimated with a binary tree partition of the predictor space. Every terminal node in the tree represents a (local) generalized GARCH model. Moreover, the GTS model allows us to exploit the additional information included in the term structure and in other relevant macroeconomic variables in a very simple way. As a consequence, the description and the interpretation of the different regimes become more intuitive. In particular, we find that the estimated regimes can be associated appropriately with particular events like the 1979-82 Fed experiment and/or business cycle contractions and expansions as determined by the Business Cycle Dating Committee of the National Bureau of Economic Research. The empirical results show that all of these generalizations are statistically and economically significant.

Comparing the forecasting accuracy of the GTS model with those of other existing models for the short-term interest rate in three out-of-sample experiments, we provide strong empirical evidence of its better potential and flexibility. At this point, our attention can turn to other particular financial applications such as the valuation of interest-rate-sensitive securities and interest rate management. In this paper we do not provide multi-step ahead forecasts for the short-term interest rate process. The GTS model does not directly yield transition dynamics for the short rate process and for all related macroeconomic variables. However, multi-step ahead forecasts can be computed in a standard way using simulation and model-based bootstrap (see, for example, Audrino and Trojani, 2004). These and other related issues are currently being explored in ongoing research.

Notes

[1]The constant term w_j in the conditional variance can be omitted. This because a de facto intercept is introduced by the square-root term. That is, $\sigma_j^2 \min r_{t-1}$ forms a lower bound for the conditional variance; see also Gray (1996).

[2]Within this framework, the conditional mean and variance could have an even more general parameterization. For example, we can introduce a new parameter γ for the level effect term, such that the conditional variance depends on $\sigma^2 r_{t-1}^{2\gamma}$ with $\gamma \neq 1/2$. However, the parameterization adopted here represents a good tradeoff between flexibility and computational feasibility.

[3]The log-likelihood is always considered conditionally on Φ_1 and on some reasonable starting value $h_1(\theta)$.

[4]Regularity conditions for the consistency of pseudo maximum likelihood estimators of tree structured GARCH models are given in Audrino and Bühlmann (2001).

[5]We adopt this approach in order to avoid a computationally unfeasible exhaustive search.

[6]Standard errors for the parameter estimates are computed using a model-based bootstrap from the standardized residuals. See Efron and Tibshirani (1993) for more details.

[7]This is consistent with the results found by Lamoureux and Lastrapes (1990).

[8]Negative (positive) in this context means below (above) a given threshold.

[9]Note that the Help Wanted Advertising in Newspapers (HELP) index is commonly though to be a leading indicator for real activity.

[10]In particular, we consider the same Gray's two-regime GRS model in (6)-(9) but we extend the time-varying transition probabilities P_t and Q_t in

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

to be of the form

$$P_t = \Phi_N(c_1 + d_1 r_{t-1} + s_1 x_{t-1}^{\text{ex}})$$

and

$$Q_t = \Phi_N(c_2 + d_2 r_{t-1} + s_2 x_{t-1}^{\text{ex}}),$$

where Φ_N is the standard normal distribution and $g_{t,j}$ $(j = 1, 2)$ is the density of a Gaussian variable with conditional mean $\mu_{t,j}$ and conditional variance $h_{t,j}$. The chosen extended two-regime GRS specification is the one with the exogenous variable x^{ex} among all the possible candidates minimizing the negative conditional (pseudo) maximum likelihood.

References

Andersen, T.G., Bollerslev,T. and Lange, S. (1999). Forecasting financial market volatility: Sample frequency vis-à-vis forecast horizon. *Journal of Empirical Finance* **6**, 457-477.

Andersen, T.G., Bollerslev,T. Diebold, F.X. and Labys, P. (2001). The Distribution of Realized Exchange Rate Volatility. *Journal of American Statistical Association* **96**, 42-55.

Ang, A. and Bekaert, G. (2002), Regime switches in interest rates. *Journal of Business and Economic Statistics* **20**, 2, 163-182.

Ang, A. and Piazzesi, M. (2003), A no-arbitrage vector autoregression of the term structure dynamics with macroeconomic and latent variables. *Journal of Monetary Economics* **50**, 745-787.

Audrino, F. and Bühlmann, P. (2001), Tree-structured GARCH models. *Journal of the Royal Statistical Society, Series B*, **63**, No. 4, 727-744.

Audrino, F. and Trojani, F. (2003), Estimating and predicting multivariate volatility regimes in global stock markets. Unpublished manuscript USI, Lugano, Switzerland.

Audrino, F. and Trojani, F. (2004), Accurate Yield Curve Scenario Generation using Functional Gradient Descent. Unpublished manuscript USI, Lugano, Switzerland.

Breiman, L., Friedman, J.H., Olshen, R.A. & Stone, C.J. (1984). *Classification and Regression Trees.* Wadsworth, Belmont (CA).

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

Dewachter, H. and Lyrio, M. (2003), Macro Factors and the Term Structure of Interest Rates. Unpublished manuscript, CES, Catholic University Leuven.

Efron, B. and Tibshirani, R.J. (1993), *An Introduction to the Bootstrap.* Chapman & Hall, London.

Estrella, A. and Mishkin, F.S. (1997), The predictive power of the term structure of interest rates in Europe and the United States: implications for the European central bank. *European Economic Review* **41**, 1375-1401.

Evans, M.D. and Lewis, K.K. (1995), Do expected shifts in inflation affects estimates of the long-run Fisher relation. *Journal of Finance* **50**, 1, 225-253.

Evans, C.L. and Marshall, D.A. (1998), Monetary policy and the term structure of nominal interest rates: evidence and theory. *Carnegie-Rochester Conference Series on Public Policy* **49**, 53-111.

Garcia, R. and Perron, P. (1996), An analysis of the real interest rate under regime shifts. *The review of Economics and Statistics* **78**, 1, 111-125.

Gray, S.F. (1996), Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics* **42**, 27-62.

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 econometric analysis of nonstationary time series and the business cycle. *Econometrica* **57**, 357-384.

Hamilton, J.D. (1994), *Time Series Analysis.* Princeton University Press, New Jersey.

Hansen, A.T. and Poulsen, R. (2000), A simple regime switching term structure model. *Finance and Stochastics* **4**, 409-429.

Kalimipalli, M. and Susmel, R. (2002), Regime-switching stochastic volatility and short-term interest rates. Unpublished manuscript, University of Houston.

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

Lewis, K.K. (1991), Was there a "peso-problem" in the US term structure of interest rates: 1979-1982?. *International Economic Review* **32**, 159-173.

Pai, J. and Pedersen, H.W. (1999), Threshold Models of the Term Structure of Interest Rate. Proceedings of the 9th International AFIR Colloquium.

Rabenmanjara, R. and Zakoian, J.M. (1993), Threshold ARCH model and asymmetries in volatility. *Journal of Applied Econometrics* **8**, 31-49.

Sola, M. and Driffill, J. (1994), Testing the term structure of interest rates using a vector autoregression with regime switching. *Journal of Economic Dynamics and Control* **18**, 601-628.

Tong, H. (1978), On a threshold model. In C.H. Chen (ed.), *Pattern Recognition and Signal Processing*, Sijthoff and Noordhoff, Amsterdam.

Tong, H. (1990), *Non-linear Time Series: A Dynamical System Approach*, Oxford Science Pubblications.

Tong, H. and Lim, K.S. (1980), Threshold autoregressive, limit cycles and cyclical data. *Journal of the Royal Statistical Society, Series B* **42**, 245-292.

Vasicek, O. (1977), An equilibrium characterization of the term structure. *Journal of Financial Economics* **5**, 177-188.

Simulations performance results

		GRS models		
	Measure	Single regime	Two regimes	GTS model
Mean				
	MAE	0.0359	0.0316	0.0324
	RMSE	0.0613	0.0590	0.0589
Variance				
	MAE	0.0833	0.0714	0.0706
	RMSE	0.1682	0.1333	0.1352

Table 1: Goodness-of-fit measures for single and two-regime generalized regime-switching (GRS) models, and the generalized tree-structured (GTS) model averaged over 50 independent simulations consisting of 500 observations generated from the two-regime GRS model with parameters given in Table 3. Classical mean absolute error (MAE) and root mean squared error (RMSE) statistics are computed for the conditional mean and variance estimates obtained using the different models.

Summary statistics of data

	Central moments				Autocorrelations		
	Mean	Stdev	Skew	Kurt	Lag 1	Lag 2	Lag 3
1 mth rates	4.6573	2.1204	1.2966	5.2241	0.9556	0.9203	0.8875
1 mth changes	-0.0038	0.6095	1.0215	15.199	-0.1045	-0.0332	-0.0653
60 mth rates	6.9773	2.4603	0.9309	3.5981	0.9860	0.9696	0.9541
Spread	2.3200	1.1502	0.6745	3.6596	0.8723	0.8064	0.7391
CPI	4.2779	2.8568	1.2667	4.1458	0.9909	0.9782	0.9629
PPI	3.4385	3.7274	1.3461	4.4884	0.9843	0.9630	0.9407
HELP	85.318	22.647	0.3225	2.2329	0.9880	0.9759	0.9601
IP	3.1998	4.6552	0.8555	3.8134	0.9622	0.9015	0.8288
UE	1.3244	10.117	0.5628	4.5274	0.9764	0.9509	0.8581
S&P500	0.6939	4.2662	0.3433	4.9306	0.0093	-0.0473	0.0120

Table 2: The 1 month yield is from the Fama CRSP Treasury bill files. The 60 month yield is annual zero coupon bond yields from the Fama CRSP bond files. Spread refers to the difference between long and short-term interest rates. The inflation measures CPI and PPI refer to CPI inflation and PPI (Finished Goods) inflation, respectively. We calculate the inflation measure at time t using $\log(P_t/P_{t-12})$ where P_t is the (seasonal adjusted) inflation index. The real activity measures HELP, IP and UE refer to the Index of Help Wanted Advertising in Newspapers, the (seasonal adjusted) growth rate in industrial production and the unemployment rate, respectively. The growth rate in industrial production is calculated using $\log(I_t/I_{t-12})$ where I_t is the (seasonal adjusted) industrial production index. S&P500 refers to S&P500 monthly log-returns. The sample period is January 1960 to December 2001, for a total of 504 observations.

GRS parameter estimates

	Single regime		Regime-switching	
Parameter	Estimate	t |(p-value)	Estimate	t |(p-value)
α_1	0.1343	2.8401*	1.0154	1.2616
α_2	–	–	0.0353	0.5930
β_1	−0.0355	−3.3495*	−0.2469	−1.1625
β_2	–	–	−0.0025	−0.1539
a_1	0.1768	2.3395*	0.1652	0.2893
b_1	0.7450	7.9980*	0	0.0004
σ_1^2	0.0047	1.6045	0.1699	1.6694
a_2	–	–	0.0001	0.0075
b_2	–	–	0.7092	6.6053*
σ_2^2	–	–	0	0.0004
c_1	–	–	2.9404	1.4268
s_1	–	–	−1.2235	−1.4088
c_2	–	–	2.5028	5.9122*
s_2	–	–	−0.2157	−2.5801*
Log-likelihood	−315.2222		−278.4033	
LB_5^2	1.6787	(0.8916)	7.2655	(0.2016)
LB_{10}^2	8.3654	(0.5932)	8.2025	(0.6091)
LB_{15}^2	9.0409	(0.8754)	12.5887	(0.6340)

Table 3: Parameter estimates and related statistics for single-regime and regime-switching GRS models. The sample period is January 1960 to December 2001, for a total of 504 monthly observations. t-statistics are based on heteroskedastic-consistent standard errors. Asterisks denote significance at the 1% level. 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:

$$\Delta r_t \mid \Phi_{t-1} \sim \begin{cases} \mathrm{N}(\alpha_1 + \beta_1 r_{t-1}, a_1\varepsilon_{t-1} + b_1 h_{t-1} + \sigma_1^2 r_{t-1}) & \text{w.p.} \quad p_{t,1}, \\ \mathrm{N}(\alpha_2 + \beta_2 r_{t-1}, a_2\varepsilon_{t-1} + b_2 h_{t-1} + \sigma_2^2 r_{t-1}) & \text{w.p.} \quad 1 - p_{t,1}, \end{cases}$$

where

$$\begin{aligned}
&\varepsilon_t = \Delta r_t - [p_{t,1}\mu_{t,1} + (1-p_{t,1})\mu_{t,2}], \\
&\mu_{t,i} = \alpha_i + \beta_i r_{t-1}, \ \ i = 1,2, \\
&h_t = p_{t,1}(\mu_{t,1}^2 + h_{t,1}) + (1-p_{t,1})(\mu_{t,2}^2 + h_{t,2}) - [p_{t,1}\mu_{t,1} + (1-p_{t,1})\mu_{t,2}]^2, \\
&p_{t,1} = (1-Q_t)\left[\frac{g_{t-1,2}(1-p_{t-1,1})}{g_{t-1,1}p_{t-1,1} + g_{t-1,2}(1-p_{t-1,1})}\right] + P_t\left[\frac{g_{t-1,1}p_{t-1,1}}{g_{t-1,1}p_{t-1,1} + g_{t-1,2}(1-p_{t-1,1})}\right], \\
&g_{t,i} = f_{\mathrm{N}}(\Delta r_t \mid S_t = i), \ \ i = 1,2, \\
&P_t = \Phi_{\mathrm{N}}(c_1 + s_1 r_{t-1}), \quad Q_t = \Phi_{\mathrm{N}}(c_2 + s_2 r_{t-1}),
\end{aligned}$$

f_{N} and Φ_{N} the density and the probability functions of the standard normal distribution. In the single-regime GRS model: $\Delta r_t \mid \Phi_{t-1} \sim \mathrm{N}(\alpha_1 + \beta_1 r_{t-1}, a_1\varepsilon_{t-1} + b_1 h_{t-1} + \sigma_1^2 r_{t-1})$.

Simple GTS parameter estimates

Regime Structure	Parameter	Optimal: $k = 4$ regimes	
		Estimate	t |(p-value)
	α_1	0.0962	1.5002
	β_1	−0.0143	−0.6367
$\Delta r_{t-1} \leq 0.101$ and $r_{t-1} \leq 4.212$	a_1	0.1798	1.7386
	b_1	0.7983	7.6234*
	σ_1^2	0	0
	α_2	−0.1729	−1.8685
	β_2	0.0541	1.7546
$\Delta r_{t-1} > 0.101$ and $r_{t-1} \leq 4.212$	a_2	0	0
	b_2	0.6284	3.4960*
	σ_2^2	0.0086	1.7709
	α_3	−0.3916	−2.4001*
	β_3	0.0690	2.1912
$4.212 < r_{t-1} \leq 6.987$	a_3	0	0
	b_3	0.5407	3.2371*
	σ_3^2	0.0219	2.8111*
	α_4	0.8280	5.6407*
	β_4	−0.1198	−5.8301*
$r_{t-1} > 6.987$	a_4	0.3657	1.8788
	b_4	0.7346	3.8690*
	σ_4^2	0.0059	0.4162
Log-likelihood		−268.1804	
LB_5^2		2.4074	(0.7904)
LB_{10}^2		12.5827	(0.2479)
LB_{15}^2		14.0075	(0.5250)

Table 4: Parameter estimates, regime's structures and related statistics for the simple generalized tree-structured (GTS) model which does not use any additional information included in other term structure and macroeconomic variables for prediction. The sample period is January 1960 to December 2001, for a total of 504 monthly observations. t-statistics are based on heteroskedastic-consistent standard errors. Asterisks denote significance at the 1% level. 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 simple GTS model: $\Delta r_t \mid \Phi_{t-1} \sim \mathrm{N}(\mu_t, h_t)$,

$$\mu_t = \sum_{j=1}^{k} (\alpha_j + \beta_j r_{t-1})\, I_{[(r_{t-1}, \varepsilon_{t-1}, h_{t-1}, \Delta r_{t-1}) \in \mathcal{R}_j]},$$

$$h_t = \sum_{j=1}^{k} \left(a_j \varepsilon_{t-1}^2 + b_j h_{t-1} + \sigma_j^2 r_{t-1}\right) I_{[(r_{t-1}, \varepsilon_{t-1}, h_{t-1}, \Delta r_{t-1}) \in \mathcal{R}_j]},$$

for a partition $\mathcal{P}$ of the state space G of $\Phi_{t-1} = \{r_{t-1}, \varepsilon_{t-1}, h_{t-1}, \Delta r_{t-1}\}$

$$\mathcal{P} = \{\mathcal{R}_1, \ldots \mathcal{R}_k\},\ G = \cup_{j=1}^{k} \mathcal{R}_j,\ \mathcal{R}_i \cap \mathcal{R}_j = \emptyset\ (i \neq j).$$

Full GTS parameter estimates

Regime Structure	Parameter	Optimal: $k = 4$ regimes	
		Estimate	t |(p-value)
	α_1	0.1580	3.0072*
	β_1	−0.0484	−3.1913*
$\text{HELP}_{t-1} \leq 88.67$	a_1	0.0001	0.0169
	b_1	0.8974	38.319*
	σ_1^2	0.0014	2.6453*
	α_2	0.8065	3.2187*
$\text{HELP}_{t-1} > 88.67$,	β_2	−0.1730	−2.8788*
$\text{CPI}_{t-1} \leq 3.497$ and	a_2	0.2245	1.5142
$\Delta r_{t-1} \leq -0.078$	b_2	0.2106	1.6472
	σ_2^2	0.0020	0.6469
	α_3	0.8621	3.2236*
$\text{HELP}_{t-1} > 88.67$,	β_3	−0.2305	−3.4858*
$\text{CPI}_{t-1} \leq 3.497$ and	a_3	0.0459	0.3982
$\Delta r_{t-1} > -0.078$	b_3	0.9429	2.9701*
	σ_3^2	0.0111	1.2135
	α_4	0.7337	4.8373*
	β_4	−0.1147	−4.3400*
$\text{HELP}_{t-1} > 88.67$ and $\text{CPI}_{t-1} > 3.497$	a_4	0.1413	1.6771
	b_4	0.8104	6.3671*
	σ_4^2	0.0085	1.4227
Log-likelihood		−263.3198	
LB_5^2		4.4333	(0.4889)
LB_{10}^2		6.6646	(0.7567)
LB_{15}^2		8.0836	(0.9204)

Table 5: Parameter estimates, regime's structures and related statistics for the full generalized tree-structured (GTS) model which use also the additional information included in the term structure and in other macroeconomic variables for prediction $(\mathbf{x}_{t-1}^{\text{ex}})$. The sample period is January 1960 to December 2001, for a total of 504 monthly observations. t-statistics are based on heteroskedastic-consistent standard errors. Asterisks denote significance at the 5% level. 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 GTS model: $\Delta r_t \mid \Phi_{t-1} \sim \text{N}(\mu_t, h_t)$,

$$\mu_t = \sum_{j=1}^{k} (\alpha_j + \beta_j r_{t-1})\, I_{[(r_{t-1},\varepsilon_{t-1},h_{t-1},\mathbf{x}_{t-1}^{\text{ex}})\in\mathcal{R}_j]},$$

$$h_t = \sum_{j=1}^{k} \left(a_j \varepsilon_{t-1}^2 + b_j h_{t-1} + \sigma_j^2 r_{t-1}\right) I_{[(r_{t-1},\varepsilon_{t-1},h_{t-1},\mathbf{x}_{t-1}^{\text{ex}})\in\mathcal{R}_j]},$$

for a partition $\mathcal{P}$ of the state space G of $\Phi_{t-1} = \{r_{t-1}, \varepsilon_{t-1}, h_{t-1}, \mathbf{x}_{t-1}^{\text{ex}}\}$

$$\mathcal{P} = \{\mathcal{R}_1, \ldots \mathcal{R}_k\},\ G = \cup_{j=1}^{k} \mathcal{R}_j,\ \mathcal{R}_i \cap \mathcal{R}_j = \emptyset\ (i \neq j).$$

Out-of-sample specification tests (conditional mean)

Model	In-sample (108 obs.): January 1970-December 1978		Out-of-sample (276 obs.): January 1979-December 2001	
	RMSE	MAE	RMSE	MAE
Single-regime GRS	0.6326	0.4353	0.7706	0.4381
Two-regime GRS	0.6246	0.4266	0.7942	0.4655
Simple GTS	0.6246	0.4370	0.7886	0.4342
Extended GRS	0.6231	0.4273	0.7896	0.4621
Full GTS	0.5452	0.3875	0.7647	0.4304
Realized volatility	0.6554	0.4504	0.7901	0.4347

Model	In-sample (144 obs.): January 1970-December 1978		Out-of-sample (240 obs.): January 1979-December 2001	
	RMSE	MAE	RMSE	MAE
Single-regime GRS	1.0096	0.6328	0.5022	0.3238
Two-regime GRS	1.0023	0.6327	0.5173	0.3375
Simple GTS	0.9958	0.6301	0.5027	0.3383
Extended GRS	0.9690	0.6104	0.5086	0.3367
Full GTS	0.9671	0.5973	0.4853	0.3038
Realized volatility	1.0002	0.6199	0.5124	0.3070

Model	In-sample (288 obs.): January 1970-December 1978		Out-of-sample (96 obs.): January 1979-December 2001	
	RMSE	MAE	RMSE	MAE
Single-regime GRS	0.8262	0.5141	0.2867	0.1790
Two-regime GRS	0.8216	0.5087	0.2851	0.1773
Simple GTS	0.8148	0.5094	0.2859	0.1789
Extended GRS	0.8207	0.5086	0.2873	0.1792
Full GTS	0.8239	0.5115	0.2485	0.1735
Realized volatility	0.8214	0.5125	0.2816	0.1805

Table 6: Predictive performance measures for single, two-regime and extended two-regime generalized regime-switching (GRS) models, simple and full generalized tree-structured (GTS) models and the realized volatility AR-GARCH model estimated on daily data. The table reports root mean squared errors (RMSE) and mean absolute errors (MAE) of the innovations $\varepsilon_t = \Delta r_t - \mu_t$. Parameters are estimated over the in-sample period, and held fixed over the out-of-sample period. The data are monthly observations of one-month Treasury bill yields. The sample extends from January 1970 to December 2001, for a total of 384 observations.

Out-of-sample specification tests (volatility)

In-sample					Out-of-sample				
Period (No. obs.)	Statistic	Single-regime GRS	Two-regime GRS	Simple GTS	Period (No. obs.)	Statistic	Single-regime GRS	Two-regime GRS	Simple GTS
January 1970-	RMSE	1.060	1.311	0.897	January 1979-	RMSE	1.951	2.579	1.916
December 1978	MAE	0.493	0.548	0.394	December 2001	MAE	0.706	0.938	0.713
(108)	R^2	0.065	0.079	0.257	(276)	R^2	0.197	−0.215	0.225
						ONL	201.7	228.6	204.4
January 1970-	RMSE	2.785	2.719	2.365	January 1982-	RMSE	0.914	1.164	0.797
December 1981	MAE	1.288	1.234	1.087	December 2001	MAE	0.389	0.486	0.419
(144)	R^2	0.109	0.348	0.269	(240)	R^2	−0.133	−0.728	0.071
						ONL	137.0	136.7	136.6
January 1970-	RMSE	1.947	1.954	1.759	January 1994-	RMSE	0.231	0.223	0.220
December 1993	MAE	0.801	0.809	0.730	December 2001	MAE	0.151	0.149	0.147
(288)	R^2	0.199	0.166	0.292	(96)	R^2	−0.047	0.025	0.046
						ONL	20.30	1.557	21.25

Table 7: Predictive performance measures for single, two-regime and extended two-regime generalized regime-switching (GRS) models, simple and full generalized tree-structured (GTS) models and the realized volatility AR-GARCH model estimated on daily data. The table reports root mean squared errors (RMSE), mean absolute errors (MAE) and R^2 between actual volatility ε_t^2, where $\varepsilon_t = \Delta r_t - \mu_t$, and the conditional variance h_t. In addition, we also consider as a measure of predictive performance the out-of-sample negative likelihood (ONL). Parameters are estimated over the in-sample period, and held fixed over the out-of-sample period. Low is better for all statistics, except for the R^2. 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 the squared actual volatility over the out-of-sample period. The data are monthly observations of one-month Treasury bill yields. The sample extends from January 1970 to December 2001, for a total of 384 observations.

Out-of-sample specification tests (volatility)

In-sample					Out-of-sample				
Period (No. obs.)	Statistic	Extended GRS	Full GTS	Realized volatility	Period (No. obs.)	Statistic	Extended GRS	Full GTS	Realized volatility
January 1970-	RMSE	1.293	0.592	0.866	January 1979-	RMSE	2.523	1.923	2.540
December 1978	MAE	0.547	0.312	0.405	December 2001	MAE	0.922	0.774	0.802
(108)	R^2	0.068	0.206	0.067	(276)	R^2	−0.198	0.262	0.134
						ONL	223.7	203.8	–
January 1970-	RMSE	2.337	2.266	2.697	January 1982-	RMSE	0.861	0.696	0.811
December 1981	MAE	1.078	0.979	1.069	December 2001	MAE	0.387	0.311	0.319
(144)	R^2	0.208	0.321	0.194	(240)	R^2	−0.211	0.296	0.086
						ONL	134.4	132.9	–
January 1970-	RMSE	1.931	2.055	1.974	January 1994-	RMSE	0.220	0.192	0.212
December 1993	MAE	0.804	0.830	0.808	December 2001	MAE	0.133	0.131	0.147
(288)	R^2	0.179	0.172	0.178	(96)	R^2	0.069	0.064	0.051
						ONL	3.969	12.59	–

Table 7: (continued)

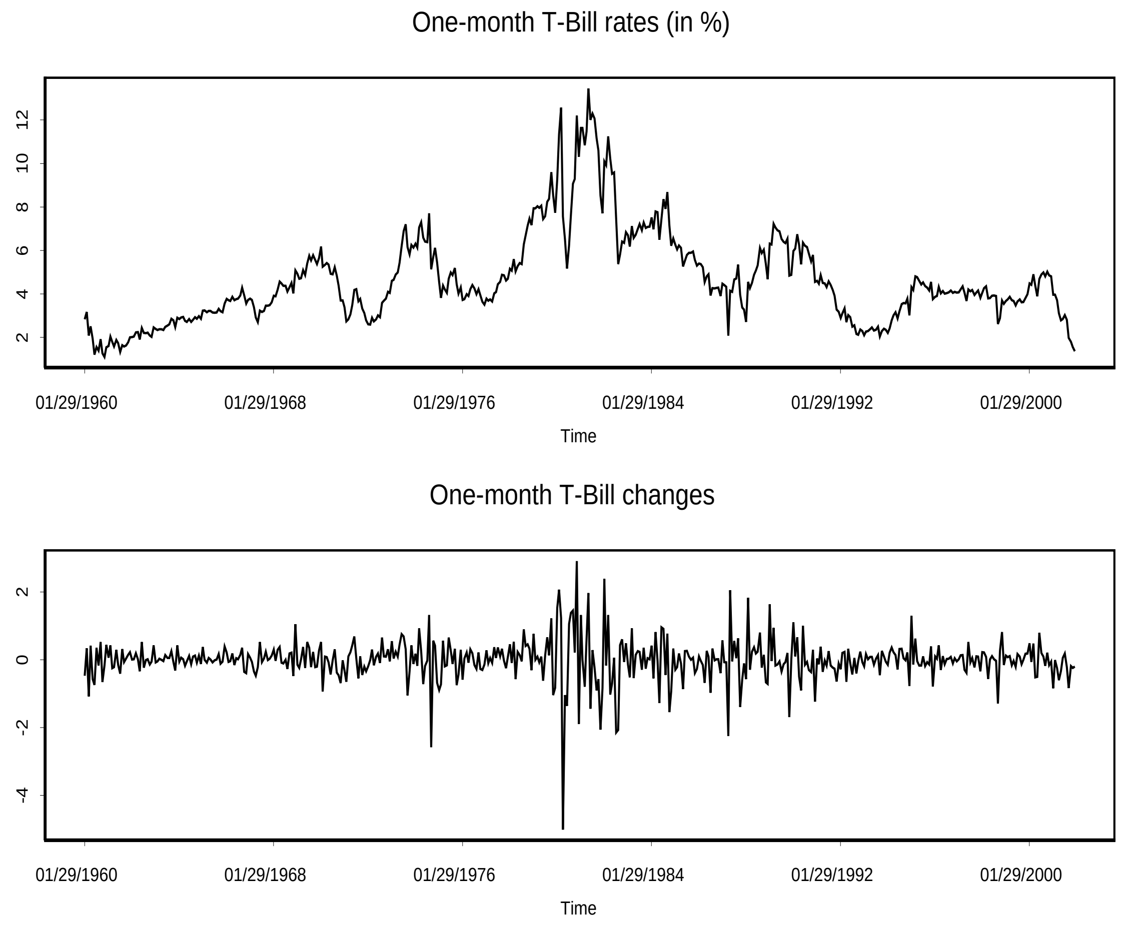

Figure 2: The top panel contains a time series of monthly one-month Treasury-bill rates (in percentages). The first differences of this series are shown in the bottom panel. The sample period is January 1960 to December 2001, for a total of 504 observations.

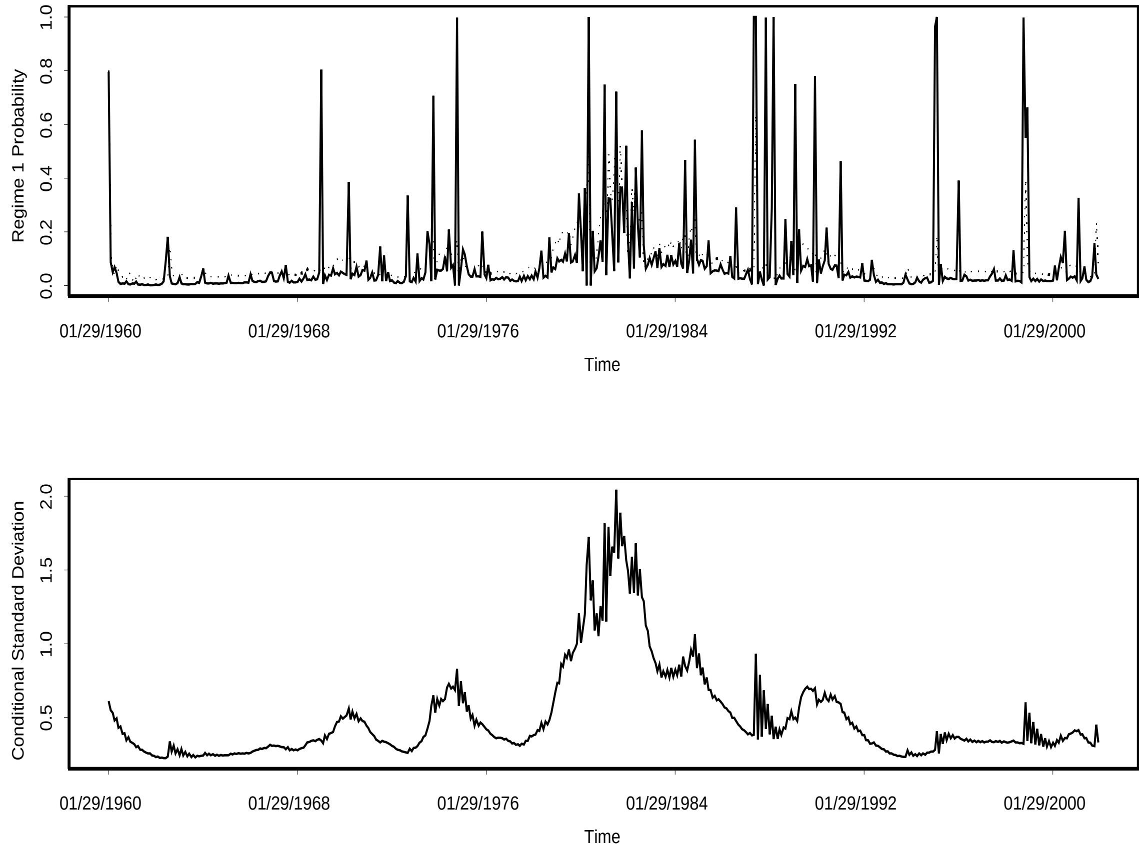

Figure 3: The top panel contains a time series plot of the ex-ante (dotted line) and smoothed (solid line) probabilities that the short rate process is in regime 1 (the high-volatility regime) at time t according to the generalized regime-switching model. The ex-ante probability is based on the information available at time t and the smoothed probability is based on the entire sample. The bottom panel contains a time series plot of the conditional standard deviation of changes in the short rate based on the generalized regime-switching model. Parameter estimates are based on a data set of one-month Treasury-bill rates. The sample period is January 1960 to December 2001, for a total of 504 monthly observations.

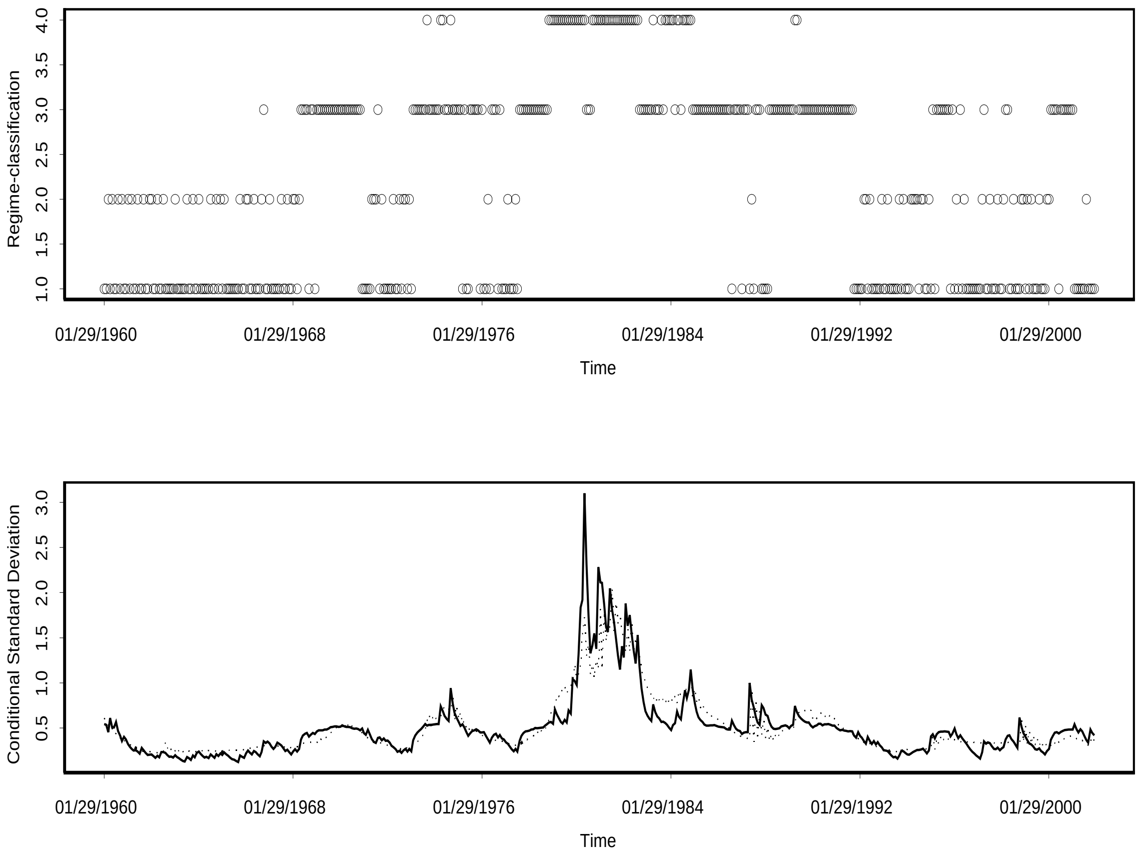

Figure 4: The top panel contains a time series plot of the regime-classification of the observation at time t (i.e. $I_t = \{1, 2, 3, 4\}$) according to the simple generalized tree-structured (GTS) model. The regime-classification is based on the endogenous information available at time $t-1$, without including additional information from the term-structure or other macroeconomic variables. The bottom panel contains a time series plot of the conditional standard deviation of changes in the short rate based on the simple generalized tree-structured (GTS) model (solid line) superimposed on those from a two-regime GRS fit (dotted line) for comparison. Parameter estimates are based on a data set of one-month Treasury-bill rates. The sample period is January 1960 to December 2001, for a total of 504 monthly observations. Optimal regimes of the simple GTS model:

$$\begin{aligned}
&\text{Regime 1:} \quad \mathcal{R}_1 = \{\Delta r_{t-1} \leq 0.101 \text{ and } r_{t-1} \leq 4.212\} \\
&\text{Regime 2:} \quad \mathcal{R}_2 = \{\Delta r_{t-1} > 0.101 \text{ and } r_{t-1} \leq 4.212\} \\
&\text{Regime 3:} \quad \mathcal{R}_3 = \{4.212 < r_{t-1} \leq 6.987\} \\
&\text{Regime 4:} \quad \mathcal{R}_4 = \{r_{t-1} > 6.987\}
\end{aligned}$$

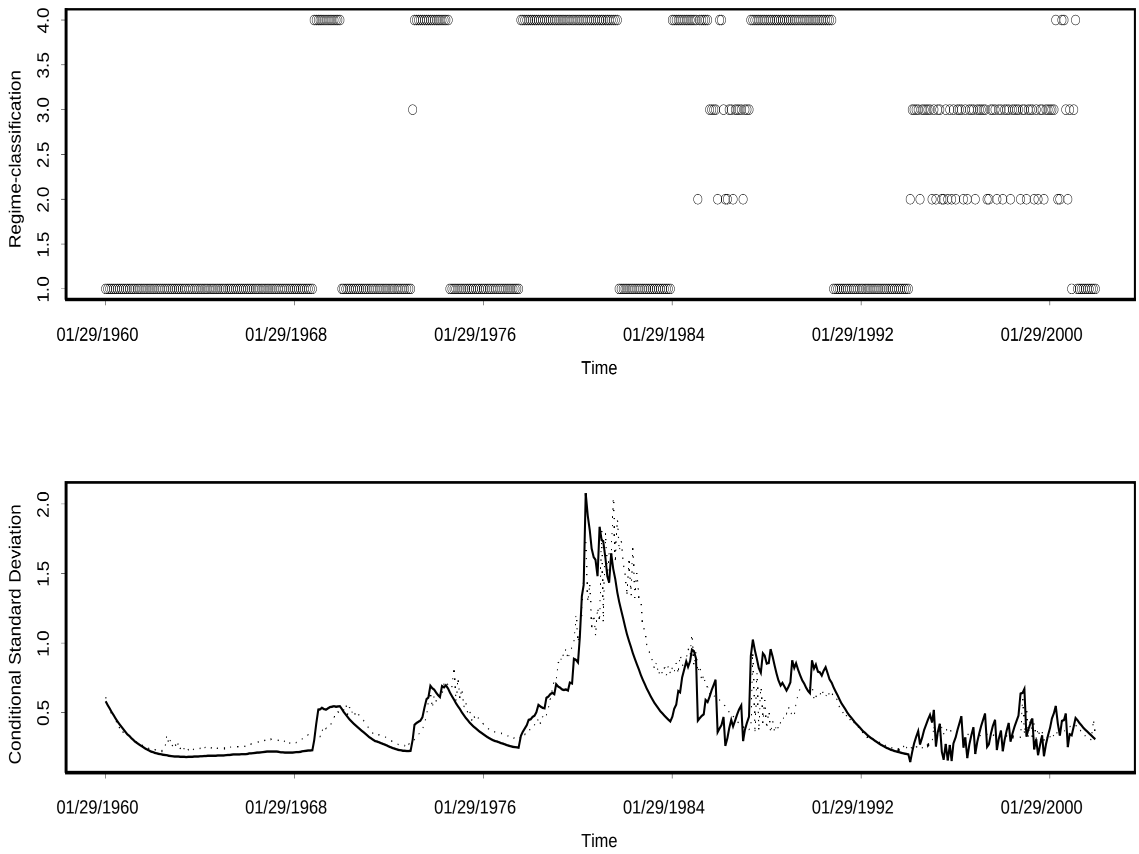

Figure 5: The top panel contains a time series plot of the regime-classification of the observation at time t (i.e. $I_t = \{1, 2, 3, 4\}$) according to the full generalized tree-structured (GTS) model. The regime-classification is based on all available information at time $t-1$, also including additional information from the term-structure and other macroeconomic variables introduced in Table 1. The bottom panel contains a time series plot of the conditional standard deviation of changes in the short rate based on the full generalized tree-structured (GTS) model (solid line) superimposed on those from a two-regime GRS fit (dotted line) for comparison. Parameter estimates are based on a data set of one-month Treasury-bill rates. The sample period is January 1960 to December 2001, for a total of 504 monthly observations. Optimal regimes of the full GTS model:

Regime 1: $\mathcal{R}_1 = \{\text{HELP}_{t-1} \leq 88.67\}$

Regime 2: $\mathcal{R}_2 = \{\text{HELP}_{t-1} > 88.67,\ \text{CPI}_{t-1} \leq 3.497 \text{ and } \Delta r_{t-1} \leq -0.078\}$

Regime 3: $\mathcal{R}_3 = \{\text{HELP}_{t-1} > 88.67,\ \text{CPI}_{t-1} \leq 3.497 \text{ and } \Delta r_{t-1} > -0.078\}$

Regime 4: $\mathcal{R}_4 = \{\text{HELP}_{t-1} > 88.67 \text{ and } \text{CPI}_{t-1} > 3.497\}$