What Drives Short Rate Dynamics? A Functional Gradient Descent Approach

Francesco Audrino

Abstract   Functional gradient descent (FGD), a recent technique coming from computational statistics, is applied to the estimation of the conditional moments of the short rate process with the goal of finding the main drivers of the drift and volatility dynamics. FGD can improve the accuracy of some reasonable starting estimates obtained using classical short rate models introduced in the literature. It exploits the predictive information of an enlarged set of variables, including yields at other maturities, time, and macroeconomic indicators. Fitting this methodology to the time series of monthly US 3-month Treasury bill rates, we find that the drift dynamics react mostly in a nonlinear way to changes in macroeconomic variables, whereas volatility dynamics are subjected to time-dependent regime-switches. Finally we show the superior performance of the final predictions obtained by applying FGD in a forecasting exercise.

Keywords   Functional gradient descent · Short rate process · Macroeconomic variables · Time-varying drift and volatility dynamics

JEL Classification   C14 · C52 · E43 · E44 · E47

1 Introduction

In the last 20 years the academic community has devoted a lot of attention to the understanding, modeling, and forecasting of the short-term interest rate process, giving rise to an enormous number of studies published in the literature. The main reason for such strong interest is the key role played by the short rate in many practical applications such as the pricing of bond and interest-rate dependent derivative

F. Audrino (✉)
Institute of Mathematics and Statistics, University of St. Gallen, Bodanstrasse 6, 9000,
St. Gallen, Switzerland
e-mail: francesco.audrino@unisg.ch

securities and (interest rate) risk management. In fact, the short rate process is generally thought to be the most important state variable driving the whole term structure of interest rates.

Recent studies in continuous time have focused on the description and estimation of (possible) non-linearities in the drift and diffusion functions of the short rate using parametric and semi- or even non-parametric approaches. Among others, it is worthwhile citing the results of Aït-Sahalia (1996b) who proposed a general parametric specification nesting most existing parametric interest rate models, in particular the classical Vasicek (1977) and Cox et al. (1985) models. Given the potential risk of model misspecification inherent in the parametric assumption yielding significant economic implications in the pricing of interest rate derivatives, nonparametric modeling has received considerable attention. Aït-Sahalia (1996a) proposed a non-parametric estimator of the diffusion function from discretely observed data in connection with a parametric drift function. Stanton (1997) proposed nonparametric estimators of the drift and diffusion functions based on different orders of approximation of the Itô process using the infinitesimal generator and Taylor series expansions. Bandi and Phillips (2003) generalized the nonparametric approach to recurrent diffusion processes, relaxing the assumption of stationarity for the short rate process. Sam and Jiang (2009) extended the nonparametric estimator proposed by Stanton (1997) by incorporating the informative potential of a panel of yields in the nonparametric estimation. They found in simulations that the proposed estimator can lead to significant efficiency gains relative to the nonparametric estimator constructed using only the time series of observed short rates, reducing problems related to spurious non-linearities in the drift function (see, for example, Pritsker 1998, or Chapman and Pearson 2000).

In the discrete setting, different time series models of the short rate dynamics have recently been proposed in order to take into account the regime-switching behavior and the heteroskedasticity and high persistence over time of the short rate process. These models also try to relate short rates with macroeconomic fundamentals like indicators of inflation and real activity. In their studies, Gray (1996), Bansal and Zhou (2002), Bansal et al. (2004), Audrino (2006), Audrino and De Giorgi (2007), and Audrino and Medeiros (2011) showed empirically that the incorporation of regime shifts (of a Markovian or threshold type) in the time series model for the short rate process significantly improves the accuracy of the estimates and predictions of short rate conditional means and variances, with important implications for the pricing of interest-rate sensitive derivative instruments. Moreover, motivated by the results illustrated in Ang and Piazzesi (2003), Diebold et al. (2006), Ang et al. (2007), and Rudebusch and Wu (2008), who considered macroeconomic fundamentals as observable factors in yield curve modeling, Audrino (2006) and Audrino and Medeiros (2011) included macroeconomic information as predictors and regime-switching variables in the construction of the short rate model, showing that such information is highly relevant for improving the fit and the prediction of the models. Indices of inflation and real activity are found to be among the main determinants, driving both the local short rate conditional dynamics in the different regimes and the way in which regime switches are determined.

This study follows the idea of extending the information set for the estimation of the first two conditional short rate moments (i.e. drift and diffusion functions) by

considering exogenous variables such as other yields and/or observable indicators of macroeconomic fundamentals as presented in Audrino (2006) and Sam and Jiang (2009). It applies a standard functional gradient descent (FGD, introduced in Audrino and Bühlmann (2003), and generalized in Audrino and Trojani (2007)) algorithm based on regression trees with the main goal of finding the most relevant forces driving the short rate dynamics among a large pool of possible candidates. FGD is strictly related to boosting, a multiple prediction and aggregation scheme for classification proposed in the 1990s by Freund and Schapire (1996): in fact, FGD yields a general representation of boosting algorithms. Starting with Breiman (1999), who showed that boosting can be viewed as an optimization algorithm in function space, many other studies have brought boosting from classification to other settings (see Bühlmann 2003, for a detailed review). In our particular case, we apply boosting to the nonlinear time series analysis of the short rate dynamics by exploiting its FGD representation.

The way FGD works is quite intuitive. FGD takes a simple parametric or non-parametric model as a first approximation and modifies it in a non-parametric way to improve a pre-specified goodness-of-fit statistic. One of the main advantages of FGD with regression trees is that it is able to choose the most relevant predictors from a large set of candidates. This helps in determining the most relevant forces driving the short rate process. A second important advantage of FGD is that the most relevant predictors for the drift function can differ from those chosen as most informative for the volatility function. This will allow us to disentangle the main forces driving the conditional mean from those determinant for the conditional variance of the short rate process. In our study, to model the short rate process we will use as starting models approaches proposed in the literature that use only the information included in the short rate time series. Applying the FGD approach we will include in the set of possible candidates yields at other maturities as well as indices of real activity and inflation.

We apply FGD to the time series of monthly US 3-month Treasury bill yields during the time period between June 1961 and July 2007. Our findings show that the additional yield curve and macroeconomic information incorporated by the FGD approach cannot be neglected and produces significantly better fits of the short rate dynamics. Interestingly, the relevant forces driving the conditional mean and variance functions of the short rate process are significantly different.

The conditional mean dynamics are driven mainly by the macroeconomic factors: they are chosen from 45 to 70% of the time, depending on the starting model. The relevance of yields at other maturities is limited, given that the relative percentage of their being chosen is always less than 25%. Surprisingly, the choice of "Time" as a predictor variable is also limited (less than 15%). Given that our starting models have no regimes, structural breaks in time do not seem to be highly relevant for modeling the time series dynamics of the short rate conditional mean in our sample period. In contrast, we find completely different results for the volatility function. In this case the most relevant predictor is always "Time." It is chosen from 56 to 75% of the times, depending on the starting model. As expected, structural breaks in time are highly relevant for modeling the short rate conditional variance dynamics and cannot be neglected. This confirms previous results found in the literature about the existence of so-called volatility regimes, such as, for example, the existence of an extremely

high volatility regime during the 1979–1982 FED experiment. Macroeconomic and term structure information seem only to be marginally relevant in this case.

The non-parametric nature of our approach can account for non-linearities of the short-rate drift and volatility functions in the dependence of all relevant predictors. In fact, our results support the estimates found for example in Sam and Jiang (2009) of a non-linear conditional mean of the short rate process.

In a final application to short rate forecasting, we show that applying FGD to the classical parametric and nonparametric estimators proposed in the literature is able to significantly improve both conditional mean and variance forecasts. In some cases, the accuracy of the final predictions obtained using simple classical models improved by FGD is similar to those one gets from very flexible approaches that take regime switches and macroeconomic information explicitly into account. The advantage of the FGD procedure is that it reduces the computational costs associated with the estimation of such flexible models of a factor of about 300. Thus, for practical reasons FGD is clearly to be preferred.

Moreover, results of a test for superior predictive ability (SPA, introduced by Hansen 2005) among the models show that FGD alone is not able to improve the out-of-sample performances of starting models that do not take into account heteroskedasticity and high persistence of short rate process over time (for example by including GARCH-type dynamics in the conditional variance equation). Such models are clearly over-performed in our forecasting exercise.

The remainder of the paper is constructed as follows. In Sect. 2 we present a general univariate nonparametric time series model of the short rate process and introduce the classical FGD procedure to estimate the conditional mean and variance functions. Empirical results of our real data investigation on monthly US 3-month Treasury bill yields are illustrated and discussed in Sect. 3. Section 4 concludes.

2 Modeling Approach

This section first introduces our univariate time series model for the conditional mean and variance functions of the short rate process. In a second step, the classical FGD (FGD) estimation procedure is reviewed, together with the algorithm that can be applied to estimate the model.

2.1 The General Model

As a starting point, let us consider a time-homogenous Itô diffusion process as a classical continuous time univariate model of the short rate r_t at time t:

$$dr_t = \mu(r_t)dt + \sigma(r_t)dw_t, \tag{1}$$

where w_t is the standard Brownian motion with $t \in [0, T]$, and $\mu(r_t)$ and $\sigma(r_t)$ are, respectively, the drift and volatility functions. In particular, most existing models of the short-term interest rate process are nested in the parametric specification introduced by Aït-Sahalia (1996b):

$$dr_t = \left(\alpha_0 + \alpha_1 r_t + \alpha_2 r_t^2 + \alpha_3 r_t^{-1}\right) dt + \sigma r_t^{\gamma} dw_t, \tag{2}$$

where $(\alpha_0, \alpha_1, \alpha_2, \alpha_3, \sigma, \gamma)$ are unknown parameters and both drift and volatility functions are specified to capture potential nonlinearities. For example, by restricting $(\alpha_2 = \alpha_3 = \gamma = 0)$ we get the Vasicek (1977) model, and by restricting $(\alpha_2 = \alpha_3 = 0, \gamma = \frac{1}{2})$ the Cox et al. (1985) model.

In this study, we focus on a discretized general version of the continuous-time model (1) of the form:

$$\Delta r_t = r_{t+\delta} - r_t = \mu_t + \varepsilon_t, \tag{3}$$

where

$$\varepsilon_t = \sqrt{h_t} z_t, \quad \mu_t = g(\Phi_{t-1}), \quad h_t = f(\Phi_{t-1}), \tag{4}$$

for some unknown conditional mean and variance functions $g(\cdot) \in \mathbb{R}$ and $f(\cdot) \in \mathbb{R}^+$, respectively, and a sampling interval $\delta > 0$. $(z_t)_{t \in \mathbb{Z}}$ is a sequence of independent identically distributed innovations with zero mean and unit variance. In model (4), the relevant conditioning information, denoted by Φ_{t-1}, is assumed to be as wide as possible. Specifically, we set $\Phi_{t-1} = \{\tilde{r}_{t-1}, \mathbf{x}_{t-1}^{ex}\}$, where $\tilde{r}_{t-1} = \{r_{t-1}, r_{t-2}, \ldots\}$ and $\mathbf{x}_{t-1}^{ex}$ is a vector of all other relevant exogenous variables used for prediction. In this study, typical factors included in $\mathbf{x}_{t-1}^{ex}$ are yields at other maturities, the spread between the long and the short rates, and some relevant macroeconomic variables such as indices for real activity and inflation already introduced in previous studies of the short rate dynamics such as Audrino (2006). 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, this model allows for a broad variety of (possibly nonlinear) shapes of the conditional mean and variance functions in reaction to past market and macroeconomic information.

It is immediate seen that the general model (3)–(4) nests the discretized version of the Aït-Sahalia (1996b) parametric continuous-time model (2), as well as discretized versions of the Vasicek (1977) and Cox et al. (1985) models as special cases. Taking the heteroskedasticity and high persistence over time features shown by the short rate process explicitly into account, one can easily specify the conditional variance function $h_t = f(\cdot)$ to include the parametric GARCH(1,1) dynamics:

$$h_t = f(\Phi_{t-1}) = f_1(\Phi_{t-1}) + w + a\varepsilon_{t-1}^2 + bh_{t-1}, \tag{5}$$

where w, a and b are the unknown parameters, and $f_1(\Phi_{t-1})$ an unspecified function depending on the whole past information set. Similarly, the model (3)–(4) nests the threshold-based regime-switching model proposed by Audrino (2006), too.

Various nonparametric estimators of the discretized drift and diffusion functions g and f have been proposed in the finance literature. In particular, we want to review here the approach proposed by Stanton (1997) and recently generalized by Sam and

Jiang (2009). Using the infinitesimal generator and Taylor series expansion, Stanton (1997) proposed nonparametric estimators of the drift and diffusion functions in (1) based on various orders of approximation of the Itô process. In greater details, Stanton (1997) proposes the following nonparametric estimators for $g(\cdot)$ and $f(\cdot)$ based on a first-order approximation of the discretized process (3)–(4):

$$\hat{g}_{\text{Stanton}}(r) = \frac{\sum_{t=1}^{T} (r_{t+1} - r_t) K_h (r_t - r)}{\sum_{t=1}^{T} K_h (r_t - r)} \quad \text{and} \tag{6}$$

$$\hat{f}_{\text{Stanton}}(r) = \frac{\sum_{t=1}^{T} (r_{t+1} - r_t)^2 K_h (r_t - r)}{\sum_{t=1}^{T} K_h (r_t - r)}, \tag{7}$$

where $K_h(u) = \frac{1}{h} K(u/h)$ and $K(\cdot)$ is a standard kernel function that satisfies usual regularity conditions. In this case, the conditioning information set Φ_{t-1} reduced to past values of the short rate process and no exogenous information is considered. The estimators (6) and (7) are obtained by setting $\delta = 1$ a fixed sampling interval. In fact, our empirical study is based on monthly data and the sampling interval δ will be equal to one month. Stanton (1997) showed with simulations in some simple settings that the error made in estimating the true drift and diffusion functions using second and third order discretized approximations based on monthly data is negligible. In our case results are qualitatively the same as those reported for the first-order approximation.

Sam and Jiang (2009) generalized Stanton (1997) approach by enlarging the information set considered in a pure endogenous estimation. They proposed a nonparametric estimator based on a panel of yields that is able to reduce some of the problems arising when using the estimators (6)–(7). In particular, Sam and Jiang (2009) showed in simulations that spurious nonlinearities or biases toward the boundaries in the estimation of the drift and volatility functions can be significantly reduced. In this study we will proceed along similar lines, enlarging the conditioning information set even further by including observable indicators of macroeconomic fundamentals in addition to yields at other maturities. The way in which the conditional mean and variance functions $g(\cdot)$ and $f(\cdot)$ are estimated using FGD is presented in the next section.

2.2 Estimation of the Conditional Mean and Variance Functions Using FGD

The main idea of FGD is to compute estimates $\hat{g}(\cdot)$ and $\hat{f}(\cdot)$ for the general, nonparametric conditional mean and variance functions $g(\cdot)$ and $f(\cdot)$ which minimize a loss function λ (generally chosen to be the negative pseudo log likelihood) under some constraints on the form of $\hat{g}(\cdot)$ and $\hat{f}(\cdot)$. These constraints are required to ensure computational feasibility, mainly in a generalized multivariate setting. More specifically, given an initial estimate $\hat{g}_0(\cdot)$ and $\hat{f}_0(\cdot)$, the estimates $\hat{g}(\cdot)$ and $\hat{f}(\cdot)$ are obtained as additive nonparametric expansions around the starting models, i.e.

$$\hat{g}(\Phi_{t-1}) = \hat{g}_0(\Phi_{t-1}) + \sum_{j=1}^{M_g} B_j^{(g)}(\Phi_{t-1}) \quad \text{and} \tag{8}$$

$$\hat{f}(\Phi_{t-1}) = \hat{f}_0(\Phi_{t-1}) + \sum_{j=1}^{M_f} B_j^{(f)}(\Phi_{t-1}), \qquad (9)$$

where the optimal stopping values M_g and M_f are generally estimated to minimize approximations of the expected prediction error. From the simple estimates of the gradient of the loss function λ, FGD determines $\hat{g}(\cdot)$ and $\hat{f}(\cdot)$ as additive nonparametric expansions of $\hat{g}_0(\cdot)$ and $\hat{f}_0(\cdot)$ which minimize the loss function λ. Therefore, FGD can be seen as a procedure that aims at producing estimates that locally improve the accuracy measured by the loss function of some initial estimates by means on nonparametric additive expansions. Such nonparametric expansions are based on some simple estimates B_j of the gradient of the loss function λ in a neighborhood of the initial estimates. These simple estimates are made using a pre-specified statistical procedure $\mathcal{B}$, called base learner. In the standard algorithm used in this study, base learners are regression trees given by:

$$B_j^{(u)}(\Phi_{t-1}) = \sum_{k=1}^{L} a_{j,u}^{(k)} I_{\left[\Phi_{t-1} \in \mathcal{R}_{j,u}^{(k)}\right]}, \quad u \in \{g, f\},$$

where $a_{j,\cdot}^{(k)}$ are some constant location parameters, L is the number of end nodes in the decision tree, and $\mathcal{R}_{j,\cdot}^{(k)}$ are the partition cells of the regression trees.

Given that regression trees have the ability to choose from among a class of predictors exactly those that are the most relevant for estimation and prediction, and given that in the classical procedures regression trees with a low number of end-nodes are considered (i.e. $L \in \{2, 3\}$), we have a perfect tool to provide an answer to our main question regarding the main forces driving the short rate dynamics. Moreover, given that the regression trees estimated in the additive expansions (8)–(9) for the conditional mean and variance functions may be different, we can disentangle the most relevant predictors driving the drift from those driving the volatility functions.

In our standard FGD procedure, we choose the loss function λ to be the negative log-likelihood implied by a "nominal" Gaussian distribution assumption for the innovations z_t in (4), i.e.

$$\lambda(r, g, f) = \frac{1}{2} \log(2\pi) + \frac{1}{2} \log(f) + \frac{1}{2f}(r - g)^2,$$

where the constant term $\frac{1}{2} \log(2\pi)$ will not affect the optimization. The optimization of λ with respect to g and f is performed by calculating the corresponding partial derivatives. This step of the optimization suggests the name *Functional Gradient Descent*.

The choice of the starting models used in the FGD algorithm is important, since FDG aims at locally improving the pseudo log likelihood criterion of an initial model estimate by means of nonparametric additive expansions. Therefore, one should start from adequate initial estimates, in order to obtain a satisfactory performance. In the real application investigated in the next section, we will consider as starting models:

1. the discretized parametric Cox et al. (1985) model;
2. a generalized version of the parametric Cox et al. (1985) model incorporating GARCH(1,1) dynamics in the conditional variance equation similar to (5) with $f_1(\Phi_{t-1}) = \sigma^2 r_t$;
3. the discretized parametric model proposed by Aït-Sahalia (1996b) given in (2);
4. the nonparametric estimators introduced by Stanton (1997) and given in (6)–(7);
5. and the threshold-based regime-switching model introduced by Audrino (2006).

Models 1, 3, and 4 are standard models introduced in the literature and used in practice. Model 2 allows us to already take heteroskedasticity and high persistence of the short rate process explicitly into account in the starting model still maintaining the computational costs associated with the whole estimation low. Model 5 is a fair competitor for the FGD improved simple models 1 to 4, in particular for the forecasting exercise performed in Sect. 3.4, already incorporating regime-shifts and macroeconomic information and yielding highly accurate out-of-sample predictions. In the case of model 5. we do not expect FGD to be able to improve significantly the predictions, given the high flexibility of the starting model. In fact it has been already shown in the past literature that it does not make much sense to use FGD in connection with too sophisticated models. The idea in this case is to verify whether FGD in connection with the other very simple, classical models can yield forecasts that are enough accurate (with respect to the benchmark model 5) and can be easily estimated, thus reducing computational time.

All details about the standard FGD algorithm used in this study can be found, for example, in Audrino and Trojani (2007).[1]

3 Empirical Results

3.1 Data

The data used in this study are 3-month U.S. Treasury bill rates downloaded on a daily basis from the Federal Reserve Board web-page and constructed as in Gürkaynak et al. (2007). In order to calculate the yields on a monthly basis, we take the last trading date of each month. We will use the daily data in our forecasting application in Sect. 3.4 to get more accurate realized measures for the unobservable conditional variances. The data span the period between June 1961 and July 2007, for a total of 554 monthly observations. Figure 1 plots the data as well as the monthly changes in short-term interest rates. Table 1 presents some sample statistics.

Figure 1 illustrates quite aptly the dramatic changes in the short-term interest rates that occurred, for example, during the FED experiment in the 1979–1982 period. The volatility of the monthly changes associated with the FED experiment is striking and stimulated the idea of modeling the short rate dynamics using different regimes for the conditional variance (see, among others, Gray 1996, or Audrino 2006). As expected and already illustrated in previous studies, Table 1 shows that the mean

---

[1] The main FGD code written in S-PLUS together with some worked out examples are available for free downloading at http://www.mathstat.unisg.ch/People.aspx.
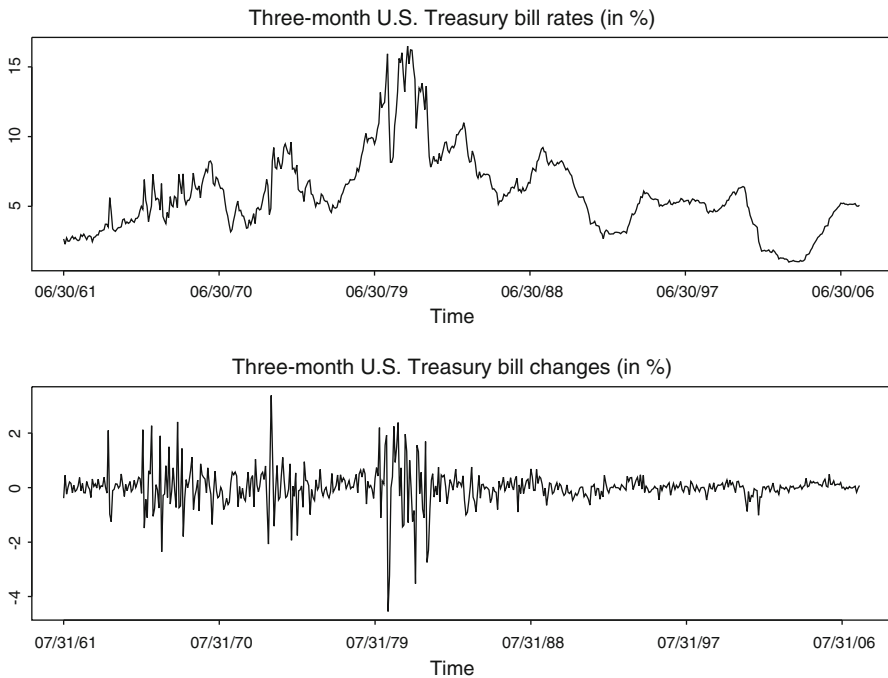
Fig. 1 The *top panel* contains a time series of monthly 3-month Treasury-bill rates (in percentages). The first differences of this series are shown in the *bottom panel*. The sample period is June 1961 to July 2007, for a total of 554 observations

change in the short-term interest rates is close to zero and that there is significant excess kurtosis.

Following the idea proposed by Sam and Jiang (2009), we download from the FED page yields at maturities 6 month, 1 year, 3 year, 5 year, and 10 year to exploit the possible additional information included in the yield curve. Some sample statistics are summarized in Table 1. As expected, the yield curve is upward-sloping.

To even further enhance the information set that we are going to use for the estimation, we also download some classical macroeconomic indicators used in the macro-finance literature to improve the accuracy of the estimation and prediction of the short rate dynamics; see, for example, Ang and Piazzesi (2003) or Audrino (2006). We consider the same (transformed) macroeconomic variables already used by Ang and Piazzesi (2003) in their study. 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). Summary statistics of these variables are reported in Table 1. In principle the FGD approach allows one to extend the prediction set by including all other variables that are thought to be relevant for the estimation and prediction of the dynamics. The algorithm itself will choose among all the most informative ones. We restrict the set to the mentioned variables because

Table 1 Summary statistics of data

	Central moments				Autocorrelations		
	Mean	SD	Skew	Kurt	Lag 1	Lag 2	Lag 3
3 month rates	5.8955	2.865	1.0757	4.802	0.973	0.9445	0.9226
3 month changes	0.0044	0.6514	−0.6817	13.3928	0.0231	−0.1221	−0.0345
6 month rates	5.9639	2.8503	0.9831	4.3958	0.9819	0.9588	0.9389
1 year rates	6.1185	2.7962	0.8948	4.0838	0.983	0.9611	0.9422
3 year rates	6.5059	2.6152	0.8855	3.746	0.9859	0.9687	0.9542
5 year rates	6.723	2.5138	0.9139	3.5923	0.9876	0.9733	0.9609
10 year rates	7.0324	2.3881	0.9308	3.4835	0.9886	0.9768	0.966
CPI	4.1763	2.7367	1.4338	4.6448	0.9908	0.9775	0.9627
PPI	3.6158	4.4418	1.0094	4.561	0.9751	0.9438	0.9138
HELP	83.75	25.1699	−0.2068	1.9566	0.9867	0.9741	0.9592
IP	3.1924	4.3061	−0.8482	4.1591	0.97	0.9219	0.8612
UE	0.9937	15.4448	1.1492	4.4364	0.9561	0.9142	0.8572

The 3 month yield is from the Federal Reserve Board web-page, as are all other yield 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. The sample period is June 1961 to July 2007, for a total of 554 observations

these have been shown to be the most important ones in several previous empirical studies.

3.2 Estimation Results

We present here the results of our FGD estimation using the different starting models introduced in Sect. 2.2 for the whole time period between June 1961 and July 2007, for a total of 554 monthly observations. Drift and volatility estimates obtained from the starting models as well as those obtained after running the FGD procedure are shown in Figs. 2, 3, 4, 5, and 6. In-sample performance measures and the optimal number of iterations in the FGD algorithm obtained using standard cross-validation are summarized in Table 2 (columns 2–4).

As expected, in the first three cases where the starting model is simple, the FGD approach is able to improve significantly the likelihood, and the optimal number of smoother that is added to the initial estimates is quite large (i.e. more than 50 up to 90, depending on the starting model). In these three cases (CIR, Aït-Sahlia, and Stanton starting models), the final drift and volatility functions are highly non-linear in the short-rate and react to different past values of a number of exogenous variables. Comparing the time series dynamics of the starting and the final estimates in Figs. 2,
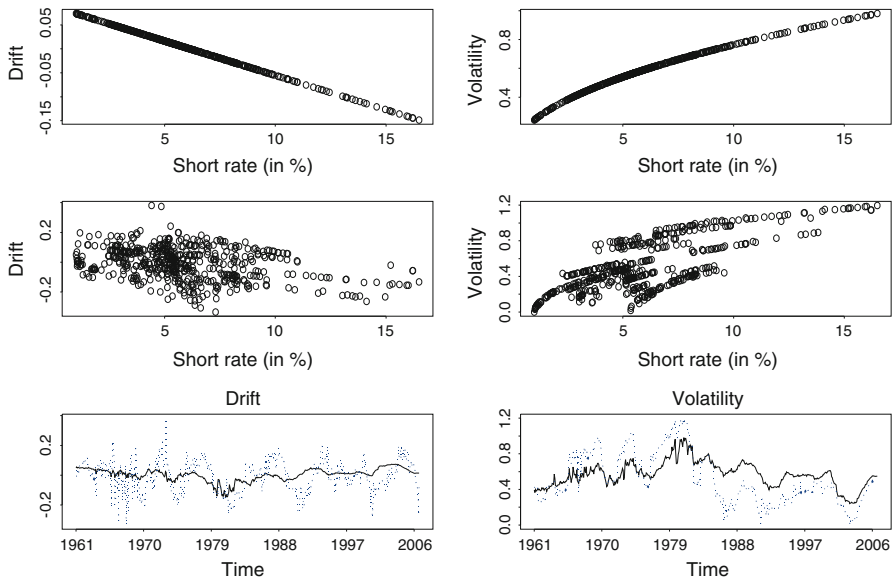
Fig. 2 The *top two panels in the first row* contain the drift and volatility functions estimated using a classical, parametric CIR model. The *two panels in the middle row* show the final drift and volatility estimates after running the FGD procedure. Finally, the *two panels in the bottom row* illustrate the differences between the estimated drift and volatility time series using the starting CIR model (*solid lines*) and the final FGD method (*dotted lines*). The sample period is June 1961 to July 2007, for a total of 554 observations

3, and 4, we see that the final drift estimates vary much more and are less smooth than the initial ones.

In the simple CIR setting (Fig. 2), differences in the volatility dynamics are also evident: one can easily recognize time periods where volatility estimates obtained from the FGD procedure are systematically larger (1965–1982 period) or smaller (after 1985) than those obtained from the CIR model. This is consistent with the empirical evidence shown in the literature about the existence of so-called volatility regimes, i.e. time periods with volatilities higher (or smaller) than the average. Such phenomena are solved in the literature using regime-switching models. The FGD procedure seems to do exactly the same by choosing "Time" as the most relevant predictor variable a considerable number of times (see the discussion below), therefore implying some positive (or negative) shocks in volatility for different time periods. Something similar also happens in the Stanton setting (Fig. 4), although not involving the very high volatility 1979–1982 period associated with the FED experiment. In contrast, no particular difference in the volatilities can be seen when the starting model is the one proposed by Aït-Sahalia (1996b) (Fig. 3).

When now considering more flexible and complex starting models like the CIR with GARCH effects in volatility or Audrino (2006) tree model that already takes into account regime-shifts in the conditional mean and volatility dynamics driven by macroeconomic variables, the FGD procedure has only a moderate (but not negligible) impact in the likelihood performance (see Table 2, fourth and fifth rows). The number
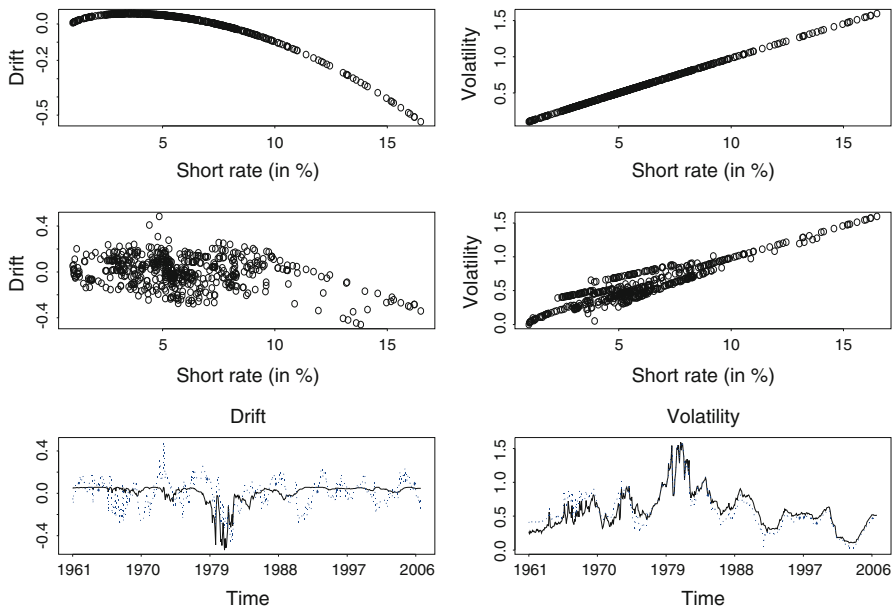
Fig. 3 The *top two panels in the first row* contain the drift and volatility functions estimated using the general parametric model proposed by Aït-Sahalia (1996b). The *two panels in the middle row* show the final drift and volatility estimates after running the FGD procedure. Finally, the *two panels in the bottom row* illustrate the differences between the estimated drift and volatility time series using the starting Aït-Sahalia (1996b) model (*solid lines*) and the final FGD method (*dotted lines*). The sample period is June 1961 to July 2007, for a total of 554 observations

of iterations in these cases is quite small (i.e. less than 10). Once again this is not surprising given the construction and purpose of FGD. As before, both drift and volatility functions are non-linear in the short rate and depend on other different explanatory variables. Similarly to the simple CIR setting, starting with the CIR-GARCH model the time-varying dynamics of the drift estimates can be significantly improved using FGD (Fig. 5). The main difference when comparing the CIR and the CIR-GARCH models as starting models yielding the initial estimates in the FGD algorithm is now that the dynamics of the volatilities are not significantly changed by the FGD procedure. Thus, the GARCH effect seems to be more relevant than possible positive (or negative) regime-shifts in time induced by the FGD algorithm.

As explained before, for comparison we also considered Audrino (2006) tree model as starting model, although we are not expecting significant changes in this case. The three local linear drift functions depending on two macroeconomic variables are clearly visible in the top left panel of Fig. 6. In this setting the FGD procedure does not seem to be able to further improve the accuracy of the conditional mean and variance dynamics. What is however important to notice is that applying the FGD methodology to a very simple starting model like the nonparametric Stanton (1997) estimators or the parametric Aït-Sahalia (1996b) model we can obtain a performance similar to the one we get using such a flexible model in a significantly reduced time: about 2 min
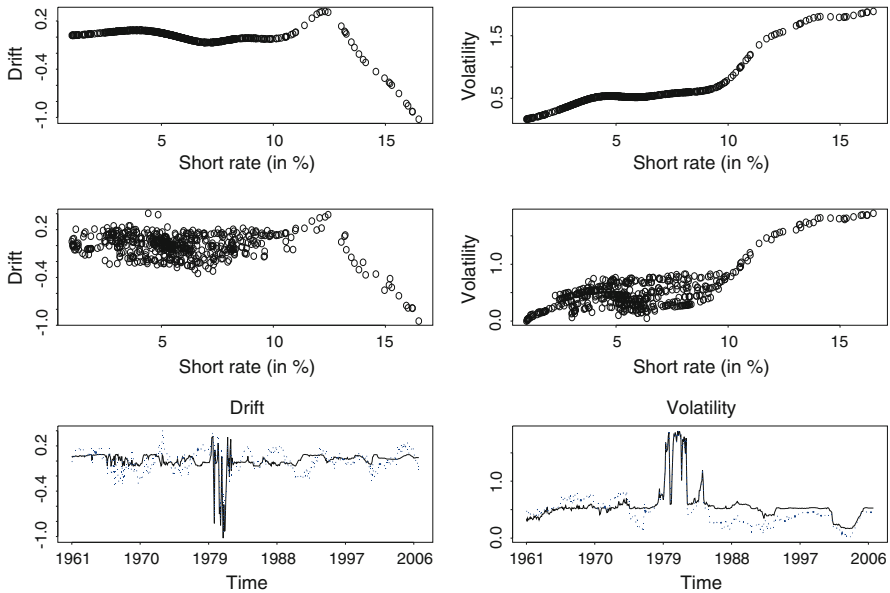
Fig. 4 The *top two panels in the first row* contain the drift and volatility functions estimated using the nonparametric approach proposed by Stanton (1997). The *two panels in the middle row* show the final drift and volatility estimates after running the FGD procedure. Finally, the *two panels in the bottom row* illustrate the differences between the estimated drift and volatility time series using the starting Stanton (1997) estimators (*solid lines*) and the final FGD method (*dotted lines*). The sample period is June 1961 to July 2007, for a total of 554 observations

versus half a day, respectively (reduction factor: about 300). This has clearly enormous consequences from a practical point of view.

3.3 Main Drivers of Drift and Volatility Functions

To end this discussion we now focus on the main question this paper poses, namely what are the main drivers of the drift and volatility dynamics of the short rate process. Results of our estimation are summarized in Table 2 (columns 5–14) and illustrated in Figs. 7 and 8 for the different starting models under investigation.

Focusing first on the drift estimates, one sees that the numbers reported in Table 2 clearly show the importance of considering macroeconomic variables as valuable predictors for the conditional mean dynamics of the short rate process. In fact, the number of times that macroeconomic indicators are chosen in the FGD procedure when the drift estimates are improved ranges from 44 to 67%, depending on the starting model. These predictors are always the most frequently chosen ones, independently of the starting model from which the initial drift estimates are computed. This is fully in line with previous empirical evidence that linked macroeconomic information with the short rate dynamics and, more generally, with the whole term structure dynamics. In looking at Fig. 7, no particular macroeconomic indicator seems to be more relevant
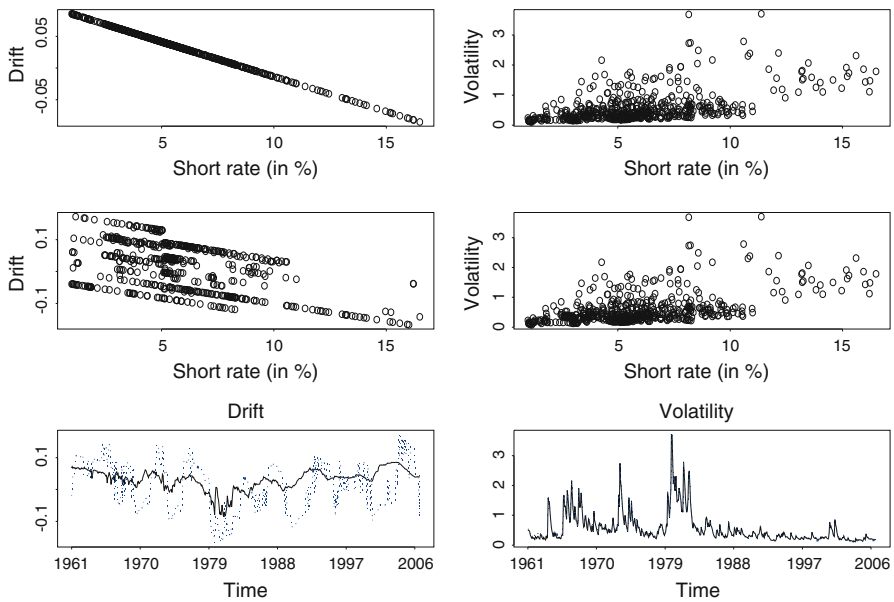
Fig. 5 The *top two panels in the first row* contain the drift and volatility functions estimated using a classical, parametric CIR-GARCH model. The *two panels in the middle row* show the final drift and volatility estimates after running the FGD procedure. Finally, the *two panels in the bottom row* illustrate the differences between the estimated drift and volatility time series using the starting CIR-GARCH model (*solid lines*) and the final FGD method (*dotted lines*). The sample period is June 1961 to July 2007, for a total of 554 observations

than the others: all variables taken into consideration in the analysis are chosen (more or less) the same number of times.

Term structure information is also quite relevant for improving the drift dynamics: yields at other maturities are chosen about 20% of the time, supporting the evidence shown in Sam and Jiang (2009). On contrast, positive (or negative) shocks in time are not found to be systematically relevant. As an implication, the regimes in time that are introduced in classical Markovian regime-switching models (see, for example, Gray 1996) seem not to be motivated directly from the need for improving the accuracy of the drift estimates.

The role of "Time" as the leading predictor for estimating the volatility dynamics is clearly highlighted by the number of times it is chosen in the FGD estimation (from 57% to more than 75%, with the sole exception of the Tree model), and is well illustrated in the pie-charts of Fig. 8. This finding supports the idea of having different short rate volatility regimes during which volatility is systematically high or low. The FGD procedure allows these regimes to be taken into account by introducing some positive (or negative) shocks in the volatility dynamics during particular time periods, in the same spirit of the regimes' construction in Markovian and threshold-based regime-switching type of models.

Macroeconomic information cannot be neglected and is exploited by the FGD algorithm about 20% of the time (depending on the starting model). Yields at other
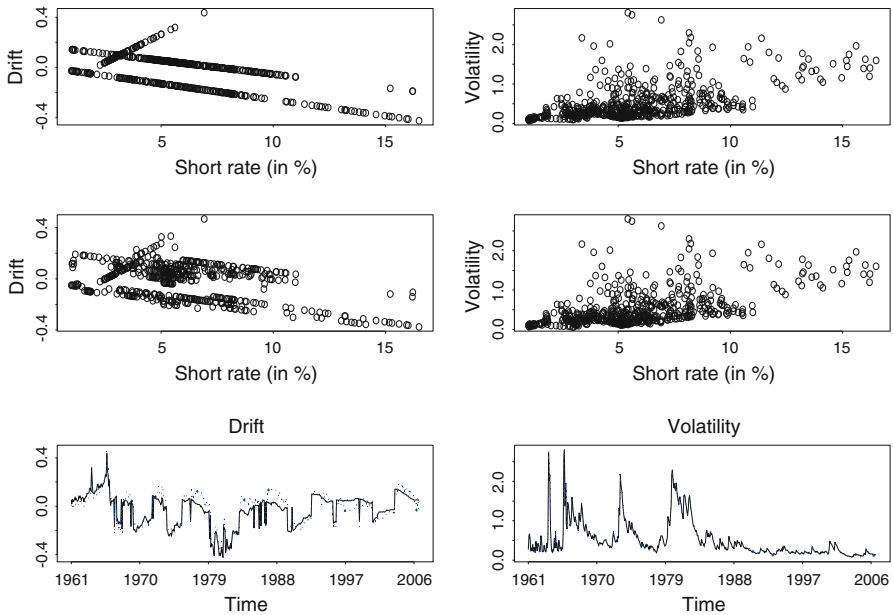
Fig. 6 The *top two panels in the first row* contain the drift and volatility functions estimated using the regime-switching tree model introduced by Audrino (2006). The *two panels in the middle row* show the final drift and volatility estimates after running the FGD procedure. Finally, the *two panels in the bottom row* illustrate the differences between the estimated drift and volatility time series using the starting Audrino (2006) model (*solid lines*) and the final FGD method (*dotted lines*). The sample period is June 1961 to July 2007, for a total of 554 observations

Table 2 Estimation results

Model	Iter.	Likelihood		Percentage of drift predictors					Percentage of volatility predictors				
		Start	End	Tot.	Time	Short rate	TS	Macro	Tot.	Time	Short rate	TS	Macro
CIR	51	443.799	255.613	51	11.2	25.9	18.5	44.4	49	56.8	0	9.1	34.1
Aït-Sahalia	91	410.769	284.991	47	4.1	13.7	23.3	58.9	53	74.5	12.7	5.5	7.3
Stanton	92	390.029	224.279	49	7.4	11.1	21.0	60.5	51	77.8	3.2	4.7	14.3
CIR-GARCH	5	317.110	271.363	60	16.7	16.7	0	66.6	40	75.0	0	0	25.0
Tree model	7	253.067	224.497	57	12.5	12.5	12.5	67.5	43	28.6	0	14.3	57.1

Results of the FGD estimation for the time period between June 1961 and July 2007, for a total of 554 observations. The first column reports on the starting models on which the FGD estimation procedure is applied. Iter. denotes the optimal number of iterations in the FGD algorithm. The likelihood as a performance measure is reported for the starting estimates (obtained using the different models) and the final estimates after running the FGD procedure. Columns 5–8 (9–14) report the percentage of the different explanatory variables chosen in the FGD estimation for the drift (volatility) function. Tot. denotes the percentage of times that the FGD procedure chose to improve the drift (or volatility) function. TS and Macro denote term structure yields different than the short rate and exogenous macro variables, respectively
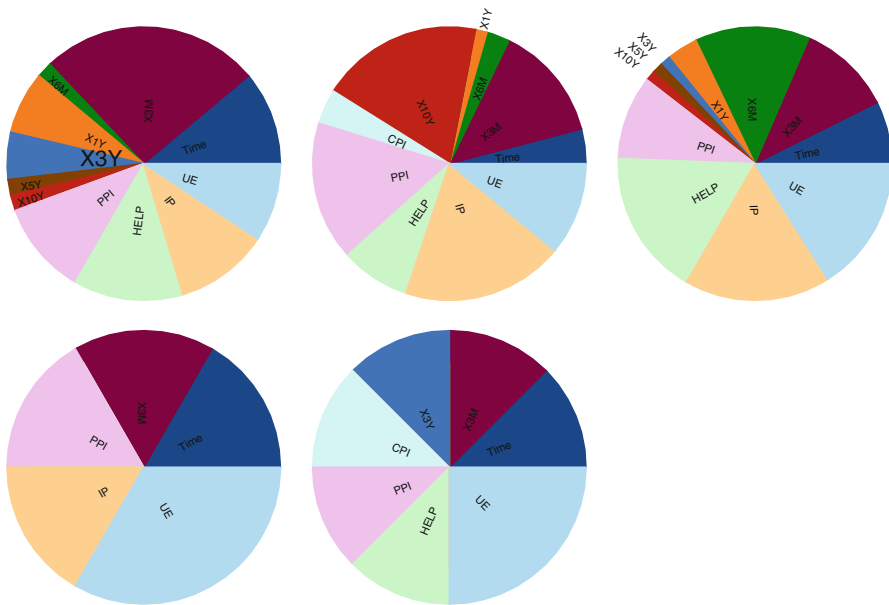
Fig. 7 Pie-charts of the predictors chosen in the FGD algorithm when improving the initial drift estimates obtained using (*first row*) the classical CIR model, the parametric Aït-Sahalia (1996b) model, the nonparametric estimator proposed by Stanton (1997), (*second row*) the CIR-GARCH model, and the threshold-based regime-switching model introduced by Audrino (2006). Possible candidates as predictors are past values of: the short rate (X3M), yields at other maturities from 7 months (X6M) to 10 years (X10Y), macroeconomic indicators for real activity (HELP, IP, UE) and inflation (CPI, PPI), and time. The sample period is June 1961 to July 2007, for a total of 554 observations

maturities do not seem to contribute very much to improving the (initial) volatility estimates of the short rate process.

3.4 Application: Forecasting the Short Rate Dynamics

In a final application, we test the predictive power of the FGD methodology starting from the estimates obtained using different approaches in a real out-of-sample exercise. To this end, we divide our sample into two parts: the in-sample estimation period runs from the beginning of the sample (June 1961) to July 2002 (494 monthly observations). The last 5 years of the sample (until July 2007) are used as an out-of-sample forecasting period. In our forecasting exercise, we re-estimate the model parameters and we re-run the FGD algorithm every month to get the forecasts of the conditional mean and variance for the next month using a standard rolling window strategy.

To quantify the goodness-of-fit of the different forecasts we use the following five criteria: the out-of-sample negative log-likelihood, the out-of-sample mean absolute error (MAE) and mean squared error (MSE) for both the conditional mean and variance forecasts. We evaluate the performance of the conditional variance forecasts with respect to realized variances computed aggregating squared differences of daily short
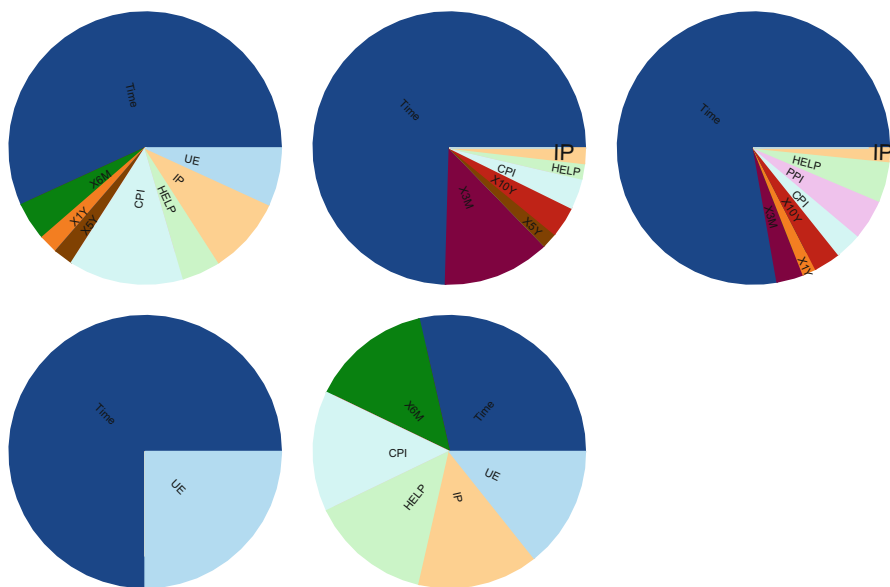
Fig. 8 Pie-charts of the predictors chosen in the FGD algorithm when improving the initial volatility estimates obtained using (*first row*) the classical CIR model, the parametric Aït-Sahalia (1996b) model, the nonparametric estimator proposed by Stanton (1997), (*second row*) the CIR-GARCH model, and the threshold-based regime-switching model introduced by Audrino (2006). Possible candidates as predictors are past values of: the short rate (X3M), yields at other maturities from 6 months (X6M) to 10 years (X10Y), macroeconomic indicators for real activity (HELP, IP, UE) and inflation (CPI, PPI), and time. The sample period is June 1961 to July 2007, for a total of 554 observations

rates (higher-frequency than monthly, about 19 to 23 daily observations each month). This should be done to overcome the problem that variances are not observable and that monthly short rate squared differences are excessively noisy estimators for the unobservable variances.

We perform a series of the SPA tests for forecasting one-month ahead first and second conditional moments introduced by Hansen (2005) to quantify statistical differences among the models. In the SPA tests, we test the null-hypothesis that each particular model is not outperformed by any of the alternative specifications. Results of our forecasting exercise are summarized in Table 3 for the same starting models introduced in Sect. 2.2. *p* Values of the SPA tests are reported in parentheses (Panel A).

It is clear from the results of the SPA tests that the best performing models for out-of-sample forecasting conditional first and second moments must explicitly take into account both non-linear regimes in the drift and volatility functions and heteroskedasticity (as modeled by the GARCH-dynamics). Overly simple models like the starting parametric CIR and Aït-Sahalia (1996b), or the non-parametric Stanton (1997) approaches are clearly overperformed by the more flexible and complex competitors. Nevertheless, in these settings the FGD procedure is always able to improve the performance and, in some cases, the FGD forecasts are not significantly less accurate than any alternative at the 5% level (this is the case, for example, of the

Table 3 Forecasting performances: SPA tests (Panel A) and Diebold and Mariano tests (Panel B)

Model	Loglik	Conditional mean		Conditional variance	
		MAE	MSE	MAE	MSE
Panel A: Forecasting performances: SPA tests					
CIR	2.0249(0)	0.1080 (0.010)	0.0205(0.027)	0.1488(0)	0.0326(0)
FGD-CIR	−6.5785(0)	0.1056(0.021)	0.0198(0.051)	0.1085(0.003)	0.0212(0.009)
Aït-Sahalia	−12.444(0)	0.1014(0.069)	0.0178(0.028)	0.1036(0.009)	0.0216(0.019)
FGD-Aït-Sahalia	−10.369(0)	0.0980(0.121)	0.01689(0.092)	0.1026(0.004)	0.0210(0.011)
Stanton	−0.9710(0)	0.1183(0)	0.0208(0.004)	0.1665(0.009)	0.0738(0.041)
FGD-Stanton	−2.0504(0)	0.1224(0)	0.0215(0.007)	0.1672(0.011)	0.0748(0.052)
CIR-GARCH	−25.551(0)	0.1059(0.052)	0.0185(0.144)	0.0331(0)	0.0033(0.558)
FGD-CIR-GARCH	−37.042(0.669)	0.0833(0.518)	0.0131(0.543)	0.0295(0)	0.0034(0.560)
Audrino's Tree	−32.768(0.296)	0.1007(0.095)	0.0170(0.143)	0.0233(0.108)	0.0032(0.860)
FGD-Audrino's Tree	−31.079(0.157)	0.0902(0.401)	0.0155(0.083)	0.0243(0.116)	0.0033(0.151)
Panel B: Forecasting performances: Diebold and Mariano tests					
CIR	−11.38(0)	−1.711(0.044)	−1.551(0.060)	−17.84(0)	−10.25(0)
Aït-Sahalia	1.541(0.062)	−1.398(0.081)	−1.190(0.117)	−1.427(0.077)	−1.749(0.040)
Stanton	−1.826(0.339)	1.154(0.124)	0.886(0.188)	0.912(0.181)	1.471(0.071)
CIR-GARCH	−4.046(0)	−3.835(0)	−2.733(0.003)	−2.361(0.009)	0.186(0.427)
Audrino's tree	0.407(0.342)	−1.260(0.104)	−0.674(0.250)	0.656(0.235)	1.452(0.101)

Panel A: Performance measures for the different models considered in the analysis. Panel B: Pairwise comparisons of the forecasts before and after running the FGD procedure (negative values favor the FGD forecasts). Loglik refers to the out-of-sample negative log-likelihood, and MAE and MSE are the mean absolute and squared errors. p Values of superior predictive ability (SPA) tests (Panel A) and pairwise generalized Diebold and Mariano tests (Panel B) are reported in parentheses

Aït-Sahalia drift forecasts improved using FGD). This is even clearer in the example of the CIR-GARCH model: in this case applying the FGD algorithm yields a significant improvement in the out-of-sample accuracy of the predictions that are then similar to those of the more complex regime-switching model.

To end the analysis, we also perform a series of generalized Diebold and Mariano tests taking into account serial correlation (see Diebold and Mariano 1995) to quantify the additional value of the FGD procedure in improving the starting forecasts of different models. Towards this goal, we perform pairwise comparisons of the performances of the starting model forecasts against those obtained after running the FGD procedure. Results are shown in Panel B of Table 3. Negative values of the statistic are in favor of the final forecasts improved using the FGD algorithm. In the parametric CIR, CIR-GARCH, and Aït-Sahalia (1996b) settings, the better forecasting power of the predictions improved using the FGD procedure is particularly evident. No statistically significant differences are found in the out-of-sample performances when starting using the non-parametric estimators introduced by Stanton (1997). This result is due to the poor performance of the starting forecasts (see Panel A of Table 3). As has been already stated in the literature, the FGD procedure is able to improve only reasonable starting values. In fact, in the cases where the initial estimates and forecasts are not at all accurate, the classical FGD technique based on regression trees is not able to produce significant improvements.

As expected the results of the tests show that improving the forecast with FGD is not necessary in cases where the starting model already takes into account the whole exogenous information included in the term structure and in other macroeconomic variables, regime-shifts and heteroskedasticity such as the model proposed by Audrino (2006). More interesting is that accurate predictions can be obtained using FGD in connection with simple models, too, allowing for a significant reduction of the computational costs.

4 Conclusions

In this study we propose the use of FGD to estimate and accurately forecast the time-varying conditional drift and volatility dynamics of the short rate process. As one of the nice features of FGD, it is able to improve any reasonable (and simple) starting estimates exploiting the information included in all predictor variables thought to be relevant in describing short rate conditional moment behavior. In fact, we allow yields at other maturities, macroeconomic indicators, and time to be chosen as predictors in the FGD procedure, with the goal of using information from the whole term structure and the macroeconomy as well as regime shifts in time to improve the initial short rate estimates.

The proposed technique can improve the estimates of some classical starting models, such as the parametric CIR and CIR-GARCH models or the non-parametric estimators proposed by Stanton (1997), in a significant way. We find that conditional drift dynamics mainly react to changes in the macroeconomy, whereas conditional volatility dynamics are mainly driven by time-dependent regime-shifts. Moreover, the heteroskedasticity of the short rate process cannot be neglected accurate estimates and

predictions of future short rate volatility are to be obtained. All these relevant features must be taken into account when introducing models for the short rate process and, more generally, models for the term structure dynamics, because of the important implications in many practical applications such as the pricing of bonds and other interest-rate dependent (derivative) securities. The construction of a joint model for the short rate dynamics and that of some relevant macroeconomic variables like inflation and real activity allowing for regime-switches (depending on time or other threshold variables) in a no-arbitrage setting is left for future research.

References

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

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

Ang, A., & 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.

Ang, A., Dong, S., & Piazzesi, M. (2007). No-arbitrage Taylor rules. Working Paper, University of Chicago.

Audrino, F. (2006). Tree-structured multiple regimes in interest rates. *Journal of Business & Economic Statistics, 24*(3), 338–353.

Audrino, F., & Bühlmann, P. (2003). Volatility estimation with functional gradient descent for very high-dimensional financial time series. *Journal of Computational Finance, 6*(3), 65–89.

Audrino, F., & De Giorgi, E. (2007). Beta regimes for the yield curve. *Journal of Financial Econometrics, 5*(3), 456–490.

Audrino, F., & Medeiros, M. C. (2011). Modeling and forecasting short-term interest rates: The benefits of smooth regimes, macroeconomic variables, and bagging. *Journal of Applied Econometrics, 26*, 999–1022.

Audrino, F., & Trojani, F. (2007). Accurate short-term yield curve forecasting using functional gradient descent. *Journal of Financial Econometrics, 5*, 591–623.

Bandi, F. M., & Phillips, P. C. B. (2003). Fully nonparametric estimation of scalar diffusion models. *Econometrica, 71*, 241–283.

Bansal, R., & Zhou, H. (2002). Term structure of interest rates with regime shifts. *Journal of Finance, 57*(5), 1997–2043.

Bansal, R., Tauchen, G., & Zhou, H. (2004). Regime shifts, risk premiums in the term structure, and the business cycle. *Journal of Business and Economic Statistics, 22*(4), 396–409.

Breiman, L. (1999). Prediction games & arcing algorithms. *Neural Computation, 11*, 1493–1517.

Bühlmann, P. (2003). Boosting methods: Why they can be useful for high-dimensional data. In *Proceedings of the 3rd international workshop on distributed statistical computing (DSC 2003)*, Vienna, Austria.

Chapman, D., & Pearson, N. (2000). Is the short rate drift actually nonlinear?. *Journal of Finance, 55*, 355–388.

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

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics, 13*, 253–263.

Diebold, F. X., Rudebusch, G. D., & Aruoba, S. B. (2006). The macroeconomy and the yield curve: A dynamic latent factor approach. *Journal of Econometrics, 131*, 309–338.

Freund, Y., & Schapire, R. E. (1996). Experiments with a new boosting algorithm. In *Machine learning: Proceedings of the thirteenth international conference*, pp. 148–156. San Francisco: Morgan Kauffman.

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

Gürkaynak, R. S., Sack, B., & Wright, J. H. (2007). The U.S. Treasury yield curve: 1961 to present. *Journal of Monetary Economics, 54*, 2291–2304.

Hansen, P. R. (2005). A test for superior predictive ability. *Journal of Business & Economic Statistics, 23*, 365–380.

Pritsker, M. (1998). Nonparametric density estimation and tests of continuous time interest rate models. *Review of Financial Studies, 11*, 449–487.

Rudebusch, G. D., & Wu, T. (2008). A macro-finance model of the term structure, monetary policy, and the economy. *The Economic Journal, 118*, 906–926.

Sam, A. G., & Jiang, G. J. (2009). Nonparametric estimation of the short rate diffusion process from a panel of yields. *Journal of Financial and Quantitative Analysis, 44*(5), 1197–1230.

Stanton, R. (1997). A nonparametric model of term structure dynamics and the market price of interest rate risk. *Journal of Finance, 52*, 1973–2002.

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