What drives short rate dynamics? A functional gradient descent approach

Audrino, Francesco
University of St. Gallen, Department of Economics
Bodanstrasse 6
9000 St. Gallen,Switzerland
E-mail: francesco.audrino@unisg.ch

In the last decades the academic community has devoted a lot of attention to the understanding, modeling, and forecasting of the short-term interest rate process, given its central role in many practical applications such as the pricing of bond and interest-rate dependent derivative securities and (interest rate) risk management.

In one approach, 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. 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. 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.

On the other hand, different time series models of the short rate dynamics in the discrete setting 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; see, among others, Audrino (2006) and Audrino and Medeiros (2010). In these studies, the researchers showed that such information is highly relevant for improving the fit and the prediction of the short rate. 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. 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 1990's. In our study, we apply boosting to the nonlinear time series analysis of the short rate dynamics by exploiting its FGD representation. To model the short rate process we will use as starting models in the FGD procedure approaches proposed in the literature that use only the information included in the short rate time series. Afterwards, 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: In fact, conditional mean dynamics are driven mainly by the macroeconomic factors. In contrast, the most relevant predictor is always "Time" for volatility, indicating that structural breaks in time are highly relevant for modeling the short rate conditional variance dynamics and cannot be neglected.

Modeling approach

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:

$$(1) \quad dr_t = \mu(r_t)dt + \sigma(r_t)dw_t,$$

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):

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

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 $(\alpha_2 = \alpha_3 = 0, \gamma = \frac{1}{2})$ we get 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:

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

where

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

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 $\Phi_{t-1} = \{\tilde{r}_{t-1}, \mathbf{x}_{t-1}^{\text{ex}}\}$, where $\tilde{r}_{t-1} = \{r_{t-1}, r_{t-2}, \ldots\}$ and $\mathbf{x}_{t-1}^{\text{ex}}$ is a vector of all other relevant exogenous variables used for prediction. In this study, typical factors included in $\mathbf{x}_{t-1}^{\text{ex}}$ 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). 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:

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

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 based on various orders of approximation of the Itô process. In greater details, he proposes the following nonparametric estimators for $g(\cdot)$ and $f(\cdot)$ based on a first-order approximation of the discretized process (3)-(4):

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

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

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.

Sam and Jiang (2009) generalized Stanton's 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 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.

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.

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

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

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. 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:

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

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, we have a perfect tool to provide an answer to our main question regarding the main forces driving the short rate dynamics.

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: (i) the discretized parametric Cox et al. (1985) model; (ii) 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 $f_1(\Phi_{t-1}) = \sigma^2 r_t$; (iii) the discretized parametric model proposed by Aït-Sahalia (1996b) given in (2); (iv) the nonparametric estimators introduced by Stanton (1997) and given in (6)-(7); (v) 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. Model 5. is a fair competitor already incorporating regime-shifts and macroeconomic information and yielding highly accurate estimates.

All details about the standard FGD algorithm used in this study can be found, for example, in Audrino and Trojani (2007). 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/org/ms/web.nsf/wwwPubInhalteGer/Francesco+Audrino?opendocument`

Estimation results

The data used in this study are three-month U.S. Treasury bill rates downloaded on a daily basis from the Federal Reserve Board web-page. In order to calculate the yields on a monthly basis, we take the last trading date of each month. The data span the period between June 1961 and July 2007, for a total of 554 monthly observations.

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. Moreover, 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, Audrino (2006). 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).

    Results of the FGD estimation obtained using standard cross-validation are reported in Table 1 and, in the case of the Aït-Sahalia (1996b) starting model, illustrated in Figure 1.

Model	Iter.	Likelihood		Percentage of Drift Predictors					Percentage of Volatility Predictors				
		Start	End	Tot.	Time	SR	TS	Macro	Tot.	Time	SR	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

Table 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 to 8 (9 to 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. SR, TS, and Macro denote short rate, term structure yields different than the short rate, and exogenous macro variables, respectively.

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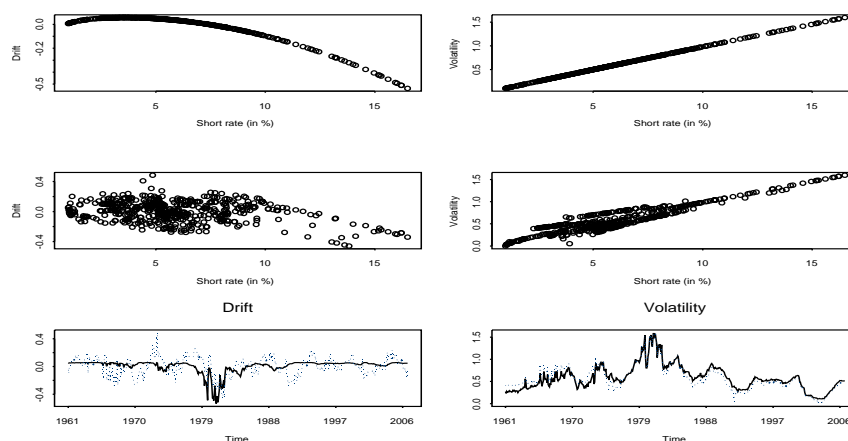

Figure 1: The top two panels in the first row contain the drift and volatility functions estimated using the general parametric model proposed by Aït-Sahalia (1996b). The two panels in the middle row show the final drift and volatility estimates after running the FGD procedure. Finally, the two panels 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.

    When now considering more flexible and complex starting models like the CIR with GARCH

effects in volatility or Audrino's 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. The number of iterations in these cases is quite small (i.e. less than 10). As before, both drift and volatility functions are non-linear in the short rate and depend on other different explanatory variables.

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 1 (columns 5 to 14). Focusing first on the drift estimates, one sees that the numbers reported in Table 1 clearly show the importance of considering macroeconomic variables as valuable predictors for the conditional mean dynamics of the short rate process. 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. No particular macroeconomic indicator seems to be more relevant than the others. 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 do not seem 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. 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 maturities do not seem to contribute very much to improving the (initial) volatility estimates of the short rate process.

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.

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

Audrino, F. and Bühlmann, P. (2003). Volatility estimation with Functional Gradient Descent for very high-dimensional financial time series. *Journal of Computational Finance* **6**, No. 3, 65-89.

Audrino, F. and Medeiros, M.C. (2010). Modeling and Forecasting short-term Interest Rates: The Benefits of Smooth Regimes, Macroeconomic Variables, and Bagging. Forthcoming in *Journal of Applied Econometrics.*

Audrino, F. and Trojani, F. (2007). Accurate Short-Term Yield Curve Forecasting using Functional Gradient Descent. *Journal of Financial Econometrics 5*, 591–623.

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

Sam, A.G. and 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.