Universität St.Gallen

Accurate Short-Term Yield Curve Forecasting using Functional Gradient Descent

Francesco Audrino a,b,_ and Fabio Trojani b,†
a Institute of Finance, University of Lugano, Switzerland
b Department of Economics, University of St. Gallen, Switzerland

Department of Economics   University of St. Gallen

Accurate Short-Term Yield Curve

Forecasting using Functional Gradient Descent

Francesco Audrino a,b,_ and Fabio Trojani b,†

Author's address: Prof. Dr. Francesco Audrino
                           USI, Institute of Finance, Via Buffi 13, CH-6900 Lugano,
                           Switzerland. E-mail: francesco.audrino@lu.unisi.ch.
                           Prof. Dr. Fabio Trojani
                           University of St. Gallen, Swiss Institute of Banking and Finance,
                           Rosenbergstr. 52, CH-9000 St. Gallen, Switzerland.
                           Tel.   +41 71 224 70 74
                           Fax   +41 71 224 70 88
                           Email   fabio.trojani@unig.ch
                           Website www.sbf.unisg.ch

Abstract

We propose a multivariate nonparametric technique for generating reliable shortterm historical yield curve scenarios and confidence intervals. The approach is based on a Functional Gradient Descent (FGD) estimation of the conditional mean vector and covariance matrix of a multivariate interest rate series. It is computationally feasible in large dimensions and it can account for non-linearities in the dependence of interest rates at all available maturities. Based on FGD we apply filtered historical simulation to compute reliable out-of-sample yield curve scenarios and confidence intervals. We back-test our methodology on daily USD bond data for forecasting horizons from 1 to 10 days. Based on several statistical performance measures we find significant evidence of a higher predictive power of our method when compared to scenarios generating techniques based on (i) factor analysis, (ii) a multivariate CCC-GARCH model, or (iii) an exponential smoothing covariances estimator as in the RiskMetricsTM approach.

Keywords

Conditional mean and variance estimation; Filtered Historical Simulation; Functional Gradient Descent; Term structure; Multivariate CCC-GARCH models.

JEL Classification

Introduction

The quality and the effectiveness of interest rate risk management depends on the ability to generate relevant forward looking yield curve scenarios that properly represent the future. Based on such scenarios, future distributions of interest rate dependent portfolio exposures and associated risk measures like VaR or Expected Shortfall can be ultimately derived from the future distributions of the underlying interest rates. In this paper we propose a new procedure to construct reliable out of sample yield curve scenarios and interval estimates for interest rate short-term risk management purposes.

One broadly used approach to the estimation of interest rate scenarios and associated risk measures is based on the historical/Monte Carlo simulation of the standardized residuals in a yield curve model with state dependent conditional means and volatilities; see for instance Barone-Adesi et al. (1998) and Barone-Adesi et al. (1999), for an introduction to the filtered historical simulation method and Jamshidian and Zhu (1997) and Reimers and Zerbs (1999) for the Monte Carlo method applied to generating term structure scenarios. While in a pure Monte Carlo setting parametric assumptions on the conditional distribution of standardized residuals have to be introduced, the historical simulation method is nonparametric and can incorporate a quite broad variety of historical distributional patterns. Since we do not want to rely on parametric assumptions on the distribution of interest rates we apply in the paper this last method to compute out-of-sample interest rate scenarios.

A necessary ingredient of the filtered historical simulation method is a dynamic model for conditional means and/or volatilities of the joint interest rate dynamics. Conditioned on an estimate of the model parameters, standardized interest rate residuals can be bootstrapped to generate out-of-sample scenarios for interest rates at different maturities. From such scenarios, confidence intervals for the prices of interest rate dependent securities can be easily derived.

The estimation of a dynamic model for the joint interest rate dynamics is a challenging task because term structures are typically high dimensional objects. Moreover, in many relevant applications it can be necessary to model not only the term structure dynamics

3

but also those of further important risk factors like for instance exchange rates. To overcome this problem, several authors have proposed some form of dimension reduction technique to reduce the estimation problem to an acceptable dimension. Examples of such methodologies are presented and discussed, among others, in Engle et al. (1990), Loretan (1997), Rodrigues (1997) and Alexander (2001). An even simpler approach to the problem is adopted by the RiskMetrics[TM] approach, which applies a multivariate exponential smoothing variance estimator to estimate conditional variances.

This paper proposes a multivariate nonparametric technique based on Functional Gradient Descent (FGD, Audrino and Bühlmann, 2003) to generate out-of-sample historical yield curve scenarios. FGD is strictly related with boosting (a multiple prediction and aggregation scheme for classification proposed in the 90's by Freund and Schapire, 1996): in fact, FGD yields a general representation of boosting algorithms. Starting with Breiman (1999), who has shown that boosting can be viewed as an optimization algorithm in function space, many other studies brought boosting from classification to other settings (see Bühlmann, 2003, for a detailed review). In our particular case, we apply boosting to multivariate nonlinear time series analysis by exploiting its FGD representation.

One advantage of FGD is that it is able to improve exactly at those individual components where some initial estimates are poorest. FGD takes a simple parametric model as a first approximation and modifies it in a non-parametric way to improve a pre-specified goodness-of-fit statistic. This strategy is similar to the SNP approach proposed by Gallant and Tauchen (1989) and extensively refined since then. The main difference between our and the SNP approach is that FGD is computationally feasible in large dimensions and allows us to estimate jointly the whole term structure dynamics, from the very short maturity segments (i.e. the overnight maturity) up to its very long end (i.e. 10 to 30 years maturity rates).[1] In contrast, the estimation of the SNP model would have serious troubles in dealing with a large cross section of yields and would have to resort to some type of variance reduction technique. In our study, we also compare the results provided by FGD with those of a model based on a three factor analysis of the yield curve dynamics. Our findings show that the additional yield curve information incorporated by the FGD approach can not be neglected and produces significantly better results, at

4

least for the application considered in this paper. Moreover, the non parametric nature of our approach can account for non-linearities in the dependence of interest rates at all available maturities. As we show below, this last feature is important in order to produce satisfactory one day ahead forecasts for interest rates in the short maturity spectrum and to predict accurately longer term maturities interest rates over horizons longer than one day.

Based on the estimated FGD dynamics we apply filtered historical simulation to USD bond data and compute short-term out-of-sample yield curve scenarios for horizons from one to ten days. We back-test the out-of-sample accuracy of our method and compare it to the one of filtered historical simulation techniques based on (i) a factor analysis of the yield curve dynamics, (ii) a multivariate AR-CCC-GARCH (Bollerslev, 1990) model and (iii) a multivariate exponential smoothing variance estimator as in the RiskMetrics$^{\text{TM}}$ approach. In particular, in (i) we estimate a three-factor model for interest rates, where the factor dynamics are driven by a 3-dimensional multivariate GARCH process extending the Diebold and Li (2004) approach in order to account for time-varying factor volatilities.

Based on several out-of-sample performance measures and formal statistical tests, we find empirical evidence of a higher predictive potential of FGD-based scenarios generating techniques. More specifically, we observe that approaches based on factor analysis or an exponential smoothing covariance matrix estimator deliver very inaccurate interval forecasts, both with respect to the expected number of back-test exceedances and the expected durations between consecutive exceedances. This finding should not, however, be interpreted as a generic disdain for factor models. Such models are at the core of modern financial theory and are well-known to be very useful, for instance for medium-term prediction of yield curve movements using macroeconomic factor variables; see, among others, Ang and Piazzesi (2003) and Diebold et al. (2005). Our results indicate, however, that structural simplifications implied by a factor analysis allowing for conditional heteroskedasticity might be too strong to predict accurately short-term yield curve movements with filtered historical simulation or related resampling procedures.

The improvement of FGD upon the CCC-based approach is smaller than the one for the factor analysis and the RiskMetrics$^{\text{TM}}$ approaches, but it is statistically signifi-

cant. Furthermore, the FGD-approach provides out of sample confidence intervals with a smoother behavior over time. On the contrary, the CCC-GARCH based confidence intervals can be often very unstable over time, especially when estimating the out-of-sample interest rates quantiles at the confidence levels typically used in risk management applications.

The paper is organizes as follows. Section 1 presents the basic model and the FGD estimation procedure needed to estimate it. A short description of the filtered historical simulation procedure is also included. Section 2 presents our application to daily USD yield curve data and the results of our back-tests. Section 3 concludes and summarizes.

1   The yield curve scenarios generating methodology

This section introduces first our multivariate model for the conditional mean and volatilities of the joint yield curve dynamics. In a second step, the FGD estimation procedure is presented, together with a computationally feasible algorithm that can be applied to estimate the model. Finally, the filtered historical simulation approach relevant for our setting is briefly reviewed.

1.1   The general model

We consider a multivariate time series $\mathcal{R} = \{\mathbf{r}_t\}_{t \in \mathbb{Z}}$, $\mathbf{r}_t = (r_{t,t+T_1}, .., r_{t+T_d})'$, of spot interest rates for a given set of fixed times to maturity $T_1 < \ldots < T_d$. Therefore, $\mathbf{r}_t$ is the yield curve at time t. Denote by $\mathcal{X} = \{\mathbf{x}_t\}_{t \in \mathbb{Z}}$, $\mathbf{x}_t = \mathbf{r}_t - \mathbf{r}_{t-1}$, the corresponding time series of interest rate changes. It is assumed that $\mathcal{R}$ is a strongly stationary process.[2] Denoting by $\mathcal{F}_{t-1}$ the information available up to time $t-1$, we model the dynamics of the conditional mean $\mu_t = \mathbf{E}(\mathbf{x}_t | \mathcal{F}_{t-1})$ and the conditional variance $V_t = \mathrm{Cov}(\mathbf{x}_t | \mathcal{F}_{t-1})$ by modeling explicitly the joint yield curve dynamics for all available maturities. No dimension reduction technique is used in the whole procedure. The basic idea is to extend the classical constant conditional correlation (CCC)–GARCH model firstly introduced by Bollerslev (1990) in order to take into account possible nonparametric nonlinearities in the functional dependence of μ_t and V_t on variables in the conditioning information set

$\mathcal{F}_{t-1}$.

We start from a time series process of the form

$$\mathbf{x}_t = \mu_t + \Sigma_t \mathbf{z}_t, \tag{1.1}$$

under the following assumptions:

(A1) (Innovations) $\{\mathbf{z}_t\}_{t \in \mathbb{Z}}$ is a sequence of i.i.d. multivariate innovations with zero mean and covariance matrix $\text{Cov}(\mathbf{z}_t | \mathcal{F}_{t-1}) = I_d$.

(A2) (CCC construction) The conditional covariance matrix $V_t = \Sigma_t \Sigma_t'$ is almost surely positive definite for all t. A typical element of V_t is given as

$$v_{t,ij} = \rho_{t,ij} (v_{t,ii} v_{t,jj})^{1/2} \quad,$$

where $i, j = 1, .., d$. The parameter $\rho_{t,ij} = \text{Corr}(x_{t,T_i}, x_{t,T_j} | \mathcal{F}_{t-1})$ is the conditional correlation between the coordinates i and j of the process $\mathcal{X}$. It is assumed in the sequel that $\rho_{t,ij}$ is constant over time: $\rho_{t,ij} = \rho_{ij}$ for some scalars $-1 \le \rho_{ij} \le 1$. Recall that by construction we have $\rho_{ii} = 1$.

(A3) (Functional form for conditional variances) The conditional variances have a non-parametric functional form given by

$$v_{t,ii} = \sigma_{t,i}^2 = \text{Var}(x_{t,T_i} | \mathcal{F}_{t-1}) = F_i(\{r_{t-j,T_k} : \ j = 1, 2, \ldots; k = 1, \ldots, d\})$$

where F_i is a function taking values in $\mathbb{R}^+$.

(A4) (Functional form for conditional means) The conditional mean μ_t has a nonparametric functional form given by

$$\mu_t = (\mu_{t,1}, \ldots, \mu_{t,d})',$$
$$\mu_{t,i} = G_i(\{r_{t-j,T_k} : \ j = 1, 2, \ldots; k = 1, \ldots, d\})$$

where G_i is a function taking values in $\mathbb{R}$.

Assumption (A1) is standard, for instance when working with multivariate time series models of the GARCH family. For estimation purposes a specific pseudo log likelihood

for $\mathbf{z}_t$ (for instance a multivariate normal one) is introduced; see Section 1.2 below. Under Assumption (A2), the conditional covariance matrix V_t of model (1.1) is of the form

$$V_t = \Sigma_t \Sigma_t^{'} = D_t R D_t,$$

where $D_t = \mathrm{diag}(\sigma_{t,1}, \ldots, \sigma_{t,d})$ and $R = [\rho_{ij}]_{i,j=1}^d$ is the matrix consisting of all (constant) correlations between the coordinates of the process $\mathcal{X}$. The nonparametric functional forms (A3)-(A4) permit a rich specification of conditional means, variances and (indirectly) conditional covariances. For instance, cross-dependencies across the different interest rates can be modeled. Similarly, a mean reversion or a nonlinearity in conditional means can be easily accounted for, as well as functional forms for conditional volatilities that are dependent on the level of current and past interest rates. Several models in the literature are included in the above setting. For instance, the standard parametric CCC-GARCH model is nested by (1.1). Similarly, multivariate AR-CCC-GARCH models where conditional means $\mu_{t,i}$ incorporate mean reversion in the standard way are special cases of the above setting. Finally, also multivariate CCC-GARCH–type models with asymmetric volatilities are nested by the above specification.

By Assumption (A2) model (1.1) avoids an explicit time varying conditional correlations structure. Models with time varying correlations have been recently advocated by Engle and Sheppard (2001), among others, in a parametric Dynamic Conditional Correlations (DCC) multivariate GARCH setting. In contrast to DCC–type models, the dynamics (1.1) is based through (A3), (A4) on a nonparametric functional form for conditional means and volatilities. This feature of the model can already account for quite flexible (nonparametric) structures in the associated time varying conditional covariances. This is an important distinction of model (1.1) from parametric DCC-GARCH-type models, which assume a more restrictive parametric GARCH-type dynamics for conditional variances and correlations. In our back-testing exercise on real data, we estimate all models using a rolling window. As shown, e.g., in Audrino and Barone-Adesi (2005) in an application to market risk measurement for stocks, this empirical approach in connection with FGD takes into account in a simple and effective way a possibly time-varying correlation structure: Qualitatively, a classical AR-DCC-GARCH estimation and a FGD-CCC-GARCH estimation yield similar results from the perspective of correlations forecasting.

8

As an additional check, we perform in Section 2.1 Tse (2000) test for the constant correlation hypothesis in the FGD residuals. The results of this test suggest the presence of a modest time variation of these conditional correlations, which can be sufficiently well taken into account for forecasting purposes by our rolling estimation procedure.

A nonparametric estimation of multivariate models of the form (1.1) in full generality is an unfeasible task, because of the course of dimensionality problem which arises when the dimension d is not a very low one. A computationally tractable, but still very general, version of the nonparametric model (1.1) can be formulated and estimated within the Functional Gradient Descent (FGD) framework (Friedman et al., 2000, and Friedman, 2001). This methodology is a general representation of classical boosting algorithms. Since Breiman (1999) result that boosting can be viewed as an optimization algorithm in function space, the FGD representation of boosting has been applied also to settings different from classification. Applications to the estimation of multivariate equity dynamics (see Audrino and Barone-Adesi, 2005 and Audrino and Bühlmann, 2003) have demonstrated that boosting via FGD is a powerful methodology which allows to construct accurate estimates for the multivariate conditional mean and covariance matrix functions also in very large dimensional applications.

In this paper, we apply the FGD technique to estimate the joint yield curve dynamics, from the very short maturity segments (i.e. the overnight maturity) up to its very long end (i.e. 10 to 30 years maturity rates). Unlike several studies on the estimation and the prediction of the yield curve, this approach avoids relying on dimension reduction techniques like Principal Components or Factor Analysis (PCA and FA, respectively). We feel that in the context of short-term yield curve scenarios generation for risk management our approach has several advantages. First, we do not need to rely on restrictive assumptions necessary to apply consistently PCA or FA in a general time series context based on stochastic conditional means and volatilities (see for instance Mardia, 1971, for an exposition of PCA and FA). Second, we can estimate the joint yield curve dynamics also over its very short term maturity spectrum, where the high variability of short-term interest rates can make the application of dimension reduction techniques cumbersome. Third, the joint term structure dynamics estimated by FGD are directly interpretable

in terms of observable interest rate variables and can be naturally related to the prices of further interest rates derivatives, as for instance forward rates. By contrast, in many applications of PCA or FA the estimated factors are typically interpreted ex post as some abstract shift-, slope- or curvature factors in the spot yield curve. These factors cannot be however naturally reconverted into forward rate factors without introducing implicitly strong restrictions in the estimated forward curve dynamics (see for instance Lekkos, 2000, for a discussion of this point). Finally, our full multivariate FGD approach allows us to account also for (possibly nonlinear) short-term feed-back effects of shocks in the very short end of the yield curve on conditional means and variances of medium and long term interest rates.

Compared to FA, the FGD approach delivers estimated dynamics that might be less economically interpretable, especially when factors can be naturally identified with some macroeconomic fundamental variables. However, our focus in this paper is on providing accurate short-term yield curve predictions for risk management, rather than on estimating an economically interpretable model of the yield curve.

The next section introduces the FGD modeling approach in a version of model (1.1) under the Assumptions (A1)–(A4).

1.2   Conditional mean and variance estimation using FGD

The main idea of FGD is to compute estimates $\widehat{G}_i(\cdot)$ and $\widehat{F}_i(\cdot)$ for the nonparametric functions $G_i(\cdot)$ and $F_i(\cdot)$, $i = 1, .., d$, which minimize a joint negative pseudo log likelihood λ under some constraints on the form of $\widehat{G}_i(\cdot)$ and $\widehat{F}_i(\cdot)$. More specifically, given an initial estimate $\widehat{G}_{i0}(\cdot)$ and $\widehat{F}_{i0}(\cdot)$, $i = 1, .., d$ – computed for instance from a parametric AR-CCC-GARCH model – estimates $\widehat{G}_i(\cdot)$ and $\widehat{F}_i(\cdot)$ are obtained as additive nonparametric expansions around $\widehat{G}_{i0}(\cdot)$ and $\widehat{F}_{i0}(\cdot)$. Such nonparametric expansions are based on some simple estimates of the gradient of the loss function λ in a neighborhood of the initial estimates $\widehat{G}_{i0}(\cdot)$ and $\widehat{F}_{i0}(\cdot)$. These simple estimates are estimated using a pre-specified statistical procedure $\mathcal{S}$, called base learner. For example, the learner could be a regression tree, a projection pursuit regressor, a neural net or a B-spline basis; see also Friedman et al. (2000), Friedman (2001), Audrino and Barone-Adesi (2005), Audrino and Bühlmann

(2003) and Bühlmann and Yu (2003) for more details. How to fit the simple estimates from the data is part of the learner. In our study, we derive the estimates using least squares.

From the simple estimates of the gradient of the loss function λ, FGD determines $\widehat{G}_i(\cdot)$ and $\widehat{F}_i(\cdot)$ as additive nonparametric expansions of $\widehat{G}_{i0}(\cdot)$ and $\widehat{F}_{i0}(\cdot)$ which minimize the joint negative pseudo log likelihood λ. Therefore, FGD aims at producing estimates which improve locally the pseudo log like likelihood of some initial estimates $\widehat{G}_{i0}(\cdot)$ and $\widehat{F}_{i0}(\cdot)$ by means on nonparametric additive expansions $\widehat{G}_i(\cdot)$ and $\widehat{F}_i(\cdot)$.

In the estimation of $\widehat{G}_i(\cdot)$ and $\widehat{F}_i(\cdot)$, we restrict the dimensionality of the predictor variables to p past lags. In other words, we estimate the nonparametric additive expansions $\widehat{G}_i(\cdot)$ and $\widehat{F}_i(\cdot)$ based on the last p multivariate observations. However, the conditional variance functions still depend on the whole history of the multivariate time series, because the initial estimates $\widehat{F}_{i0}(\cdot)$ are obtained from a simple multivariate GARCH-type model.[3] The restriction of using only p lagged multivariate observations to define $\widehat{G}_i(\cdot)$ and $\widehat{F}_i(\cdot)$ is not really a strong one in our case, because typically a moderate number of lags is sufficient to model the cross-dependence between variables in the conditional mean and variance dynamics of our interest rate series. At the same time, the auto-dependence in the conditional mean and variance of our interest rate series is already taken into account by an infinite lag-polynomial in the definition of $\widehat{G}_{i0}(\cdot)$ and $\widehat{F}_{i0}(\cdot)$.

Conditionally on the first p observations, the negative pseudo log likelihood implied by a "nominal" Gaussian distribution assumption for $\mathbf{z}_t$ in (1.1) is given by:

$$
-\sum_{t=p+1}^{n} \log\left((2\pi)^{-d/2}\det(V_t)^{-1/2}\exp(-\xi_t^T V_t^{-1}\xi_t/2)\right)
$$

$$
= \sum_{t=p+1}^{n}\left(\log(\det(D_t)) + \frac{1}{2}(D_t^{-1}\xi_t)'R^{-1}(D_t^{-1}\xi_t)\right) + n'd\log(2\pi)/2 + n'\log(\det(R))/2
$$

(1.2)

where $\xi_t = \mathbf{x}_t - \mu_t$, D_t is a diagonal matrix with elements $\sqrt{v_{t,ii}}$ and $n' = n - p$. Therefore,

a natural conditional loss function for our FGD estimation procedure is

$$\lambda_R(\mathbf{x}, \mathbf{G}, \mathbf{F}) = \log(\det(D(\mathbf{F}))) + \frac{1}{2}(D(\mathbf{F})^{-1}(\mathbf{x} - \mathbf{G}))' R^{-1}(D(\mathbf{F})^{-1}(\mathbf{x} - \mathbf{G}))$$
$$+ \frac{1}{2}\log(\det(R)) + \frac{d}{2}\log(2\pi),$$
$$D(\mathbf{F}) = \text{diag}(\sqrt{F_1}, \ldots, \sqrt{F_d}),$$
$$\mathbf{x} - \mathbf{G} = (x_1 - G_1, \ldots, x_d - G_d)', \tag{1.3}$$

where the terms $d\log(2\pi)/2$ and $\log(\det(R))/2$ are constants that do not affect the optimization.

As highlighted by the subscript R, the loss function λ_R depends on the unknown constant correlation matrix R. At any step of our FGD optimization procedure, the updated optimal values of R, $\mathbf{G}$, $\mathbf{F}$ will be constructed by a two step procedure. For a given initial correlation matrix R, updated estimates for all G_i's and F_i's are obtained by minimizing λ_R with respect to $\mathbf{G}$, $\mathbf{F}$. In a second step, given the updated estimates $\widehat{\mathbf{G}}$ and $\widehat{\mathbf{F}}$ the correlation matrix is updated using the empirical moments of the resulting standardized multivariate residuals. Therefore, given estimates $\hat{\mathbf{G}} = (\widehat{G}_1, \ldots, \widehat{G}_d)$ and $\hat{\mathbf{F}} = (\hat{F}_1, \ldots, \hat{F}_d)$, we compute the standardized residuals

$$\hat{\varepsilon}_{t,i} = \left(x_{t,i} - \widehat{G}_i(\mathbf{r}_{t-1}, \ldots)\right)/\hat{F}_i(\mathbf{r}_{t-1}, \ldots)^{1/2}, \ t = p+1, \ldots, n$$

to obtain the empirical correlation matrix

$$\hat{R} = (n - p)^{-1} \sum_{t=p+1}^{n} \hat{\varepsilon}_t \hat{\varepsilon}_t^T, \ \hat{\varepsilon}_t = (\hat{\varepsilon}_{t,1}, \ldots, \hat{\varepsilon}_{t,d})', \tag{1.4}$$

as an updated estimate of R.[4]

The optimization of λ_R with respect to $\mathbf{G}$, $\mathbf{F}$ is performed by calculating the partial derivatives of the loss function λ_R with respect to all G_i's and F_i's. In our setting, they are given for any $i = 1, \ldots, d$, by

$$\frac{\partial \lambda_R(\mathbf{x}, \mathbf{G}, \mathbf{F})}{\partial G_i} = -\sum_{j=1}^{d} \frac{\gamma_{ij}(x_j - G_j)}{F_i^{1/2} F_j^{1/2}}, \tag{1.5}$$

and

$$\frac{\partial \lambda_R(\mathbf{x}, \mathbf{G}, \mathbf{F})}{\partial F_i} = \frac{1}{2}\left(\frac{1}{F_i} - \sum_{j=1}^{d} \frac{\gamma_{ij}(x_i - G_i)(x_j - G_j)}{F_i^{3/2} F_j^{1/2}}\right), \tag{1.6}$$

12

respectively, where $[\gamma_{ij}]^d_{i,j=1} = R^{-1}$. This step of the optimization suggests the name *Functional Gradient Descent*. Indeed, given initial estimates $\widehat{G}_{i0}(\cdot)$ and $\widehat{F}_{i0}(\cdot)$, $i = 1, .., d$, the above gradients are used by FGD to define a set of simple additive expansions of the functions $\widehat{G}_{i0}(\cdot)$ and $\widehat{F}_{i0}(\cdot)$. Such expansions improve the optimization criterion precisely in the directions of steepest descent of the loss function λ_R. Moreover, since they define a nonparametric estimate of **G** and **F** the resulting optimization is a functional one.

Details on the FGD algorithm used in the paper are presented below. In Step 2 of the algorithm the above gradients are fitted by means of a base learner $\mathcal{S}$. In Step 3 and 4, the estimated gradients are used to define a set of additive expansions $\widehat{G}_{i0}(\cdot)$ and $\widehat{F}_{i0}(\cdot)$ which improve the optimization criterion precisely in the directions of steepest descent of λ.

Algorithm: Estimation of the conditional means and volatilities

Step 1 (initialization). Choose appropriate starting function $\hat{G}_{i,0}(\cdot)$ and $\hat{F}_{i,0}(\cdot)$ and define for $i = 1, .., d$ and $t = p + 1, .., n$:

$$\hat{G}_{i,0}(t) = \hat{G}_{i,0}(\mathbf{r}_{t-1}, \mathbf{r}_{t-2}, \ldots)$$
$$\hat{F}_{i,0}(t) = \hat{F}_{i,0}(\mathbf{r}_{t-1}, \mathbf{r}_{t-2}, \ldots).$$

Compute $\hat{R}_0$ as in (1.4) using $\hat{\mathbf{G}}_0$ and $\hat{\mathbf{F}}_0$. Set $m = 1$. Natural starting functions in our application are univariate AR-GARCH estimates for the single components, $i = 1, \ldots, d$, of the process $\mathcal{X}$. In particular, the conditional mean of interest rate changes in the initializing estimate depends on past multivariate interest rate levels in its autoregressive structure. The GARCH structure of the chosen initializing variance functions implies FGD variance estimates that are functions of the whole process history and not of only a finite number of process lags.

Step 2 (projection of component gradients to base learner). For every component $i = 1, \ldots, d$, perform the following steps.

(I) (mean) Compute the negative gradient

$$U_{t,i} = -\frac{\partial \lambda_{\hat{R}_{m-1}}(\mathbf{x}_t, \mathbf{G}, \hat{\mathbf{F}}_{m-1}(t))}{\partial G_i}|_{\mathbf{G}=\hat{\mathbf{G}}_{m-1}(t)}, \ t = p + 1, \ldots, n.$$

This is explicitly given in (1.5). Then, fit the negative gradient vector $U_i = (U_{p+1,i}, \ldots, U_{n,i})'$ with a base learner $\mathcal{S}$, using the first p time-lagged predictor variables (i.e. $\mathbf{r}_{t-p}^{t-1} = (\mathbf{r}_{t-1}, \ldots, \mathbf{r}_{t-p})'$ is the predictor for $U_{t,i}$):

$$\hat{g}_{m,i}(\cdot) = \mathcal{S}_X(U_i)(\cdot),$$

where $\mathcal{S}_X(U_i)(x)$ denotes the predicted value at x from the base learner $\mathcal{S}$ using the response vector U_i and a predictor variable X (say). In our application, we use as predictor variables the first two lags of multivariate interest rate levels, i.e. $p = 2$. In fact, we found in our empirical analysis that adding more than two lagged multivariate interest rate levels as predictor variables does not improve the results.[5]

(II) (variance) Compute the negative gradient

$$W_{t,i} = -\frac{\partial \lambda_{\hat{R}_{m-1}}(\mathbf{x}_t, \hat{\mathbf{G}}_{m-1}(t), \mathbf{F})}{\partial F_i}\Big|_{\mathbf{F}=\hat{\mathbf{F}}_{m-1}(t)}, \quad t = p+1, \ldots, n.$$

This is explicitly given in (1.6). Then, analogously to (I) fit the negative gradient vector $W_i = (W_{p+1,i}, \ldots, W_{n,i})'$ with the base learner $\mathcal{S}$, using again the first p time-lagged predictor variables

$$\hat{f}_{m,i}(\cdot) = \mathcal{S}_X(W_i)(\cdot).$$

For the same reasons given above, we use again as predictor variables the first two lags of the multivariate interest rate levels.

Step 3 *(line search).* For every $j = 1, \ldots, d$, perform a one-dimensional optimization for the step-length,

$$\hat{w}_{m,i}^{(\mathrm{me})} = \mathrm{argmin}_w \sum_{t=p+1}^{n} \lambda_{\hat{R}_{m-1}}(\mathbf{x}_t, \hat{\mathbf{G}}_{m-1}(t) + w\hat{g}_{m,i}(\mathbf{r}_{t-p}^{t-1}), \hat{\mathbf{F}}_{m-1}(t)),$$

$$\hat{w}_{m,i}^{(\mathrm{vol})} = \mathrm{argmin}_w \sum_{t=p+1}^{n} \lambda_{\hat{R}_{m-1}}(\mathbf{x}_t, \hat{\mathbf{G}}_{m-1}(t), \hat{\mathbf{F}}_{m-1}(t) + w\hat{f}_{m,i}(\mathbf{r}_{t-p}^{t-1})),$$

where $\hat{\mathbf{G}}_{m-1}(t) + w\hat{g}_{m,i}(\cdot)$ and $\hat{\mathbf{F}}_{m-1}(t) + w\hat{f}_{m,i}(\cdot)$ are defined as the functions which are constructed by adding in the $i-$th component only.[6]

Step 4 (up-date). Select the best component $j \in \{1, \ldots, d\}$ for the conditional mean and

variance, respectively, as

$$i_m^{*(\text{me})} = \operatorname{argmin}_i \sum_{t=p+1}^{n} \lambda_{\hat{R}_{m-1}}(\mathbf{x}_t, \hat{\mathbf{G}}_{m-1}(t) + \hat{w}_{m,i}^{(\text{me})} \hat{g}_{m,i}(\mathbf{r}_{t-p}^{t-1}), \hat{\mathbf{F}}_{m-1}(t))$$

$$i_m^{*(\text{vol})} = \operatorname{argmin}_i \sum_{t=p+1}^{n} \lambda_{\hat{R}_{m-1}}(\mathbf{x}_t, \hat{\mathbf{G}}_{m-1}(t), \hat{\mathbf{F}}_{m-1}(t) + \hat{w}_{m,i}^{(\text{vol})} \hat{f}_{m,i}(\mathbf{r}_{t-p}^{t-1})).$$

If the improvement in minimizing the empirical criterion (1.2) for the component $i_m^{*(\text{me})}$ in the conditional mean is larger than the one for the component $i_m^{*(\text{vol})}$ in the conditional variance, then up-date as

$$\hat{\mathbf{G}}_m(\cdot) = \hat{\mathbf{G}}_{m-1}(\cdot) + \hat{w}_{m,i_m^{*(\text{me})}}^{(\text{me})} \hat{g}_{m,i_m^{*(\text{me})}}(\cdot),$$

$$\hat{\mathbf{F}}_m(\cdot) = \hat{\mathbf{F}}_{m-1}(\cdot)$$

and set $j_m^* = 1$. Else, up-date as

$$\hat{\mathbf{G}}_m(\cdot) = \hat{\mathbf{G}}_{m-1}(\cdot),$$

$$\hat{\mathbf{F}}_m(\cdot) = \hat{\mathbf{F}}_{m-1}(\cdot) + \hat{w}_{m,i_m^{*(\text{vol})}}^{(\text{vol})} \hat{f}_{m,i_m^{*(\text{vol})}}(\cdot)$$

and set $j_m^* = 2$. Then, compute the new estimate $\hat{R}_m$ according to (1.4) using $\hat{\mathbf{G}}_m$ and $\hat{\mathbf{F}}_m$.

Step 5 (iteration). Increase m by one and iterate Steps 2–4 up to an optimal level $m = M$. The optimal level M is determined by means of a cross-validation procedure discussed in more detail in Remark 4 below. The resulting functions $\hat{\mathbf{G}}_M$, $\hat{\mathbf{F}}_M$ are our FGD estimates for conditional means and volatilities. More formally, they are given by:

$$\hat{\mathbf{G}}_M(\cdot) = \hat{\mathbf{G}}_0(\cdot) + \sum_{m=1}^{M} \hat{w}_{m,i_m^{*(\text{me})}}^{(\text{me})} \hat{g}_{m,i_m^{*(\text{me})}}(\cdot) I_{\{j_m^*=1\}}$$

$$\hat{\mathbf{F}}_M(\cdot) = \hat{\mathbf{F}}_0(\cdot) + \sum_{m=1}^{M} \hat{w}_{m,i_m^{*(\text{vol})}}^{(\text{vol})} \hat{f}_{m,i_m^{*(\text{vol})}}(\cdot) I_{\{j_m^*=2\}}.$$

Remark 1. The base learner $\mathcal{S}$ in Step 2 determines the FGD estimates $\hat{\mathbf{G}}_M(\cdot)$ and $\hat{\mathbf{F}}_M(\cdot)$ via the predicted values of the gradients U_i and W_i, $i = 1,..,d$, implied by the objective

function λ. This base learner should be a "weak" one - not involving a too large number of parameters to be estimated - in order to avoid an immediate overfitted estimate at the first iteration of the algorithm. The complexity of the FGD estimates $\hat{\mathbf{G}}_M(\cdot)$ and $\hat{\mathbf{F}}_M(\cdot)$ is increased by adding further nonparametric terms at every step of the above iterations. In our application, these additional terms are nonparametric functions of the first two lags of multivariate interest rate levels. However, since the starting point $\widehat{F}_0(\cdot)$ of our FGD construction is a GARCH-type variance function, estimated FGD conditional variances depend on the whole process history. We use decision trees as base learners, because particularly in high dimensions they perform a very effective variable selection by exploiting only a few explanatory variables as optimal predictors. However, this is not an exclusive choice: other base learners could be applied and compared based on some form of cross-validation criterion.

When using decision trees as base learners, the additive term functions $g_{m,i}(\cdot)$ and $f_{m,i}(\cdot)$ introduced in the steps 2 and 3 of the FGD algorithm are explicitly given by:

$$g_{m,i}(\mathbf{r}_{t-p}^{t-1}) = \sum_{k=1}^{L} a_{m,i}^{(k)} I_{[\mathbf{r}_{t-p}^{t-1} \in \mathcal{R}_{m,i}^{(k)}]},$$

$$f_{m,i}(\mathbf{r}_{t-p}^{t-1}) = \sum_{k=1}^{L} b_{m,i}^{(k)} I_{[\mathbf{r}_{t-p}^{t-1} \in \mathcal{C}_{m,i}^{(k)}]},$$

where $a_{m,i}^{(k)}$ $(b_{m,i}^{(k)})$ are some constant location parameters and L is the number of end nodes in the decision tree. The cells $\mathcal{R}_{m,i}^{(k)}$ $(\mathcal{C}_{m,i}^{(k)})$ are constructed by fitting with Least Squares a regression tree to the negative gradient vectors U_i (W_i), $i = 1, .., d$, in step 2 of the FGD algorithm.

Remark 2. As mentioned, it is desirable to use sufficiently "weak" base learners in the above FGD algorithm. A simple effective way to reduce the complexity of a base learner is via shrinkage towards zero. In this case, the up-date Step 4 of the FGD algorithm can be replaced by an updating step given by:

$$\hat{\mathbf{G}}_m(\cdot) = \hat{\mathbf{G}}_{m-1}(\cdot) + \nu \cdot \hat{w}_{m,i_m^{*(\mathrm{me})}}^{(\mathrm{me})} \hat{g}_{m,i_m^{*(\mathrm{me})}}(\cdot) \text{ or }$$

$$\hat{\mathbf{F}}_m(\cdot) = \hat{\mathbf{F}}_{m-1}(\cdot) + \nu \cdot \hat{w}_{m,i_m^{*(\mathrm{vol})}}^{(\mathrm{vol})} \hat{f}_{m,i_m^{*(\mathrm{vol})}}(\cdot), \tag{1.7}$$

where $\nu \in [0,1]$ is a shrinkage factor on a fixed grid, which can be determined by a

cross-validation procedure. This approach reduces the variance of the base learner by the factor ν^2. The resulting location parameters $\nu \cdot \hat{w}_{m,i_m^{*(\text{vol})}}^{(\text{vol})} \hat{b}_{m,i_m^{*(\text{vol})}}^{(k)}$, $k = 1, \ldots, L$, of the regression tree in the additive expansion for the conditional variances are constrained to be nonnegative in order to guarantee positivity of the final FGD conditional variance estimates.

Remark 3. The initialization Step 1 in the FGD algorithm is important, since FDG aims at improving locally by means of nonparametric additive expansions the pseudo log likelihood criterion of an initial model estimate. Therefore, one should start from adequate initial estimates, in order to obtain a satisfactory performance. In our application, we make use of the fit of a diagonal VAR(p_i)-CCC-GARCH(1,1) model[7] to initialize the FGD algorithm by means of functions $G_{i,0}$, $F_{i,0}$, $i = 1, .., d$, given by

$$G_{i,0}(\mathbf{r}_{t-1}, \mathbf{r}_{t-2}, \ldots) = \mu_{t,i} = \sum_{k=1}^{p_i} \phi_{k,i} x_{t-k,T_i},$$

$$F_{i,0}(\mathbf{r}_{t-1}, \mathbf{r}_{t-2}, \ldots) = \sigma_{t,i}^2 = \alpha_{0,i} + \alpha_{1,i}(x_{t-1,T_i} - \mu_{t-1,i})^2 + \beta_i \sigma_{t-1,i}^2,$$

where the autoregressive parameter p_i is selected in order to optimize the Akaike's Information Criterion (AIC) for each individual series i. The choice of this particular criterion does not affect the final results of our analysis. We obtain similar forecasting results when using the consistent and more parsimonious Schwartz Bayesian Information criterion (SBIC). This feature is related to the fact that both criteria yield reasonable starting conditional mean functions for our FGD algorithm. These starting functions are slightly modified (if necessary) during the main FGD estimation step. In addition, in most cases we obtain the same lag parameter p for the univariate conditional mean and variance dynamics when using the AIC or the SBIC criterion.

Our initial estimates depend on the whole history of the process in the GARCH–part of the model. Therefore, the resulting FDG estimates also imply individual variance structures that depend on the whole process history. Using a Gaussian pseudo likelihood function, we estimate by pseudo maximum likelihood the initial model for each of the d individual series, thereby neglecting in the first step the structure of the correlation matrix R. This feature causes some loss in efficiency but has the advantage that the model estimation remains fast and therefore computable also in very high dimensions d.

Remark 4. The stopping criterion in Step 5 of the FGD algorithm is important. It can be viewed as a regularization device that is particularly effective when fitting a complex model. We determine the optimal step M of the algorithm by means of a cross validation scheme. Given a sample size n, we split the sample into two subsamples of sample sizes $0.7 \cdot n$ and $0.3 \cdot n$, respectively. The first subsample is used as training set, the second as test set. The optimal step M in the algorithm is chosen as the step m that optimizes the cross-validated log-likelihood based on these two subsamples. In our 12-dimensional real data application, typical values of the optimal parameter M ranged from $M = 10$ to $M = 20$, which implies a number of estimated additive expansion terms in the FGD estimates between 12 and 20.[8]

Consistency results for the above FGD estimation procedure based on convex risk minimization criteria are available; see, among others, Bühlmann (2006), Mannor et al. (2003), Zhang and Yu (2005), Zhang (2004) and Lugosi and Vayatis (2004). Most of these works consider regression or classification trees as base learners. These FGD–consistency proofs hold for additive expansions of the form (1.7), in which the terms of order $m \geq 1$ are functions defined on a fix finite-dimensional domain. This is one of the reasons why in our FGD we work with additive expansions in which the higher order terms are functions of a finite number p of lags of our multivariate interest rate series.

1.3 Simulation of future yield curve scenarios

Based on the FGD estimates for the multivariate conditional mean vector μ_t and for the covariance matrix V_t, we apply a filtered historical simulation procedure to generate out-of-sample scenarios for the term structure of interest rates. This procedure is briefly reviewed in the next section. The whole real data analysis presented in Section 2 is obtained using S-PLUS. The main FGD software together with some worked examples is available for free downloading at `http://www.people.lu.unisi.ch/audrinof`.

We generate future scenarios for the time series $\mathcal{R}$ of interest rate changes (and consequently the time series $\mathcal{X}$ of interest rate levels). To this end, we apply a multivariate version of the filtered historical simulation procedure. Our historical simulation is based on a model-based bootstrap of multivariate filtered historical residuals, implied by an FGD

estimation of the yield curve dynamics. Using the bootstrapped residuals, we construct out of sample scenarios for the term structure. The FGD model estimate is used as the filter for the estimation of standardized multivariate residuals. Since our primary purpose is short-term yield curve forecasting, we do not impose anywhere additional no-arbitrage constraints on the simulated yield curve scenarios, which would be also very difficult to impose, given the nonparametric nature of our approach.[9]

More details on the complete simulation methodology are as follows. In a first step, we filter the multivariate standardized innovations $\mathbf{z}_t$ with our model (1.1):

$$\mathbf{z}_t = (\Sigma_t)^{-1}(\mathbf{x}_t - \mu_t),$$

$$V_t = \Sigma_t \Sigma_t^T = D_t R D_t, \ t = 1, \ldots, n,$$

where the individual conditional mean functions $\mu_{t,i} = G_i(\cdot)$ and variance functions $\sigma_{t,i}^2 = F_i(\cdot)$, $i = 1, \ldots, d$ are estimated by means of our FGD technique, as described in detail by the algorithm of Section 1.2. Under Assumption (A1), the standardized multivariate innovations are i.i.d. and can be therefore bootstrapped. The historical standardized residuals are drawn randomly (with replacement) and are used to generate pathways for future interest rate changes (and, consequently, for future interest rate levels). Hence, we apply a model-based bootstrap (Efron and Tibshirani, 1993) where from an i.i.d. resampling of the standardized multivariate residuals $\mathbf{z}_t$ we recursively generate a time series of interest rates using the structure and the fitted parameters of the estimated optimal model (1.1).

Specifically, we draw randomly dates with corresponding standardized innovations

$$\mathbf{z}_1^*, \mathbf{z}_2^*, \ldots, \mathbf{z}_x^*, \tag{1.8}$$

where x is the time horizon at which we want to generate future scenarios (typically, from 1 up to 10 days). We then construct for each time to maturity T_i pathways for future conditional means and (squared) volatilities and interest rate levels, from time $n + 1$ up to time $n + x$ (say), based on the model structure (1.1). More formally we compute the

quantities

$$\widehat{\mu}^*_{t+b,i} = \widehat{G}_i(\{r^*_{t+b-s,k}; s = 1, 2, \ldots, p, \ k = 1, \ldots, d\}),$$

$$\widehat{v}^*_{t+b,ii} = (\widehat{\sigma}^*_{t+b,i})^2 = \widehat{F}_i(\{r^*_{t+b-s,k}; s = 1, 2, \ldots, p, \ k = 1, \ldots, d\}),$$

$$\widehat{v}^*_{t+b,ij} = \widehat{\rho}_{ij} \sqrt{\widehat{v}^*_{t+b,ii} \widehat{v}^*_{t+b,jj}} \ ,$$

$$x^*_{t+b,T_i} = \widehat{\mu}^*_{t+b,i} + (\widehat{\Sigma}^*_{t+b} \hat{\mathbf{z}}^*_b)_i,$$

$$r^*_{t+b,T_i} = r^*_{t+b-1,T_i} + x^*_{t+b,T_i}, \quad b = 1, \ldots, x, \ i, j = 1, \ldots, d, \tag{1.9}$$

where all quantities denoted by "$\widehat{}$" are based on the model structure estimated by means of the FGD algorithm in section 1.2.

The "empirical" distribution of simulated model-based interest rate levels at the chosen future time point $n + x$ for each series $i = 1, \ldots, d$, is obtained by replicating the above simulation procedure a large number of times, e.g. 2000 times. Note that all the parameters estimated in the main FGD step are kept fixed during the 2000 replications. Confidence bounds for the term structure of interest rates at the future time point $n + x$ for a confidence level q are finally estimated by the lower and upper $\frac{1-q}{2}$-quantiles of the simulated "empirical" distribution of interest rates. In our exposition we focus for brevity on confidence levels $q = 0.90, 0.95, 0.99$. However, any other quantile of the simulated interest rates distribution could be estimated in the same way.

All the codes for our historical simulation procedure are written in S-PLUS and are ran on a standard PC. The running time needed to get the final results is in the order of a couple of hours and the running time of the main FGD estimation is in the order of a few minutes.

2 Empirical Results

In this section we back-test on real data our FGD scenario generation technique for forecasting horizons $x = 1, 3, 5, 10$ days and for three different confidence levels $q = 0.90, 0.95, 0.99$.

We compare the performance of our approach with three historical simulation procedures based on (i) a generalization of the three factor analysis for the yield curve dy-

namics proposed in Diebold and Li (2004), (ii) the industry standard benchmark[10] used by RiskMetrics[TM] and (iii) a standard multivariate AR-CCC-GARCH model. The third comparison is particularly useful, because it highlights the exact contribution of the FGD technique in enhancing the accuracy of VaR predictions for the yield curve relatively to a standard multivariate GARCH model. All models compared, including our FGD proposal, are based on a constant conditional correlations assumption.

2.1 Data

We consider multivariate time series for the yield curves of daily interest rate levels r_{t,T_i} at twelve different maturities T_i. For the lowest maturity segments, i.e. overnight, 1 week, 2 weeks, 1 month, 2 months, 3 months, 6 months and 1 year, we make use of Euro dollar interest rates. For the higher maturities, i.e. 2 years, 5 years, 10 years and 30 years, we make use of interest rates of US government bonds. The data span the time period between January 1, 1996 and September 30, 2002, for a total of 1760 trading days, and have been downloaded from *Data Stream International*. We split our sample in a back-testing period used to test the predictive accuracy of our FGD methodology and an in-sample estimation period used to initialize the model parameter estimates. The back-testing period goes from January 3, 2000 to September 30, 2002, for a total of 716 trading days. In our back-testing exercise the model parameters are re-estimated every 20 working days, as new data become available for prediction purposes, using all multivariate past observations in the estimation of the model dynamics. Note that this choice is not restrictive. Qualitatively, the results are the same also when re-estimating parameters every week. However, we found that re-estimating the parameters every month is a good trade-off between flexibility and computational feasibility. The updated first and second moment dynamics are then used to compute out of sample VaR predictions based on historical simulation for the whole back-testing period.

Table 1 presents summary statistics of the time series of interest rate changes in our sample. Figure 1 plots the yield curves in our sample as a function of time and maturity.

TABLE 1 AND FIGURE 1 ABOUT HERE.

21

Table 1 shows that the sample means of all interest rate changes in our sample are negative, highlighting the fact that in our back-testing period the Fed reduced several times the target interest rate. This effect is more pronounced for interest rates up to 2 years times to maturity and is clearly visible in Figure 1. In particular, we can expect a back-test based on such a time span to be a quite hard test for a VaR prediction model. Finally, the volatilities for interest rates up to 1 month time to maturity tend to be larger than those of interest rates corresponding to further time to maturities. The Ljung-Box statistics LB(20) testing for autocorrelations in the level of interest rate changes up to the 20th order are strongly significant for maturities up to 1 year, showing evidence of some autocorrelation at shorter times to maturity for the euro bonds interest rates in our sample. For higher times to maturity they are not significant at the 5% confidence level.

The |LB(20)| statistics for testing the null hypothesis of no autocorrelation in the absolute interest rate changes are all highly significant, supporting a volatility clustering hypothesis. These results are not specific for the chosen order 20, but remain qualitatively the same also for other orders in the range 1 to 30. Finally, when analyzing the sample correlations between interest rates of different maturities (not reported here) we observe that, as expected, the time series of interest rate changes of different times to maturities are positively correlated, with higher correlations for the longer times to maturity; for example, the sample correlations of interest rate changes at 3 and 6 months and at 2 and 5 years are 0.73 and 0.91, respectively.

Figure 1 highlights that our interest rate data contains a moderate number of huge short-term swings in the low-maturity segment. This feature is shared also by other interest rate data sets we have investigated. In other applications, the presence of potentially large swings in the data might have to be addressed in a more systematic way, in order to avoid a too high sensitivity of the historical simulation results to these influential points. This is not an issue for the FGD step of our procedure itself, because the FGD point estimates are obtained by improving a cross-validated criterion that depends on the out-of-sample performance of the model. More caution has to be applied, however, to the estimation of the initial univariate GARCH(1,1) processes in our procedure, because the resulting point estimates could even imply, e.g., a non-stationarity of the estimated pro-

22

cess. A systematic way to circumvent the non-robustness of GARCH estimation methods based on a Gaussian pseudo maximum likelihood has been proposed in Mancini, Ronchetti and Trojani (2005), who developed robust estimation and inference procedures for these model settings. These methods could be used to estimate the underlying GARCH(1,1) dynamics in the first step of our historical simulation approach.

To end this preliminary analysis, we perform the test for constant correlations in Tse (2000) using our starting in-sample data[11] and get a value of the LMC statistic of 94.73, which implies a p-value of 0.0118. Therefore, we obtain some evidence of a moderate time variation in the conditional correlation matrix of our data. To account for a possible variation in correlations, we apply a rolling window estimation to our CCC-FGD model. This simple way of accounting for time varying correlations together with the FGD model structure is already enough to produce the clearly higher forecasting power of our model.

Starting from these summary statistics, it is reasonable to model the joint yield curve dynamics based on some multivariate GARCH-type model of the general form (1.1). We apply the FGD technique of Section 1 and investigate the accuracy of its VaR predictions for short-term horizons typically used in risk management applications. In particular, using FGD we can account for a possibly non-linear dependence between multivariate interest rate series. Moreover, we do not need any dimension reduction technique like FA. In fact, historical simulation combined with a FA of the joint yield curve dynamics provides very poor, typically too conservative, VaR predictions in our study. This holds also for forecasts of interest rates in the long term maturity spectrum. To investigate this issue we estimated a three factor model on our yield curve data and computed the implied out-of-sample VaR predictions using historical simulation. The three factor model we use is a generalization of the model recently proposed by Diebold and Li (2004) where, in addition, we allow for three-dimensional multivariate GARCH dynamics in the innovations to the level, slope and curvature. We applied such FA to several subsets of times to maturity in our sample, in order to control for the impact of the highly variable short-term interest rates on the prediction results implied by such a FA. However, we always obtained poor, typically too conservative, interest rate interval estimates.

As an illustration, Figure 2 presents the estimated one day ahead 95%-confidence intervals for the 10-years maturity interest rate when using the FGD– (dot dashed curves) and the FA–based (dotted curves) approaches. Similar findings arise for the other maturities.

FIGURE 2 ABOUT HERE.

In Figure 2 it appears clearly that the interval estimates provided by the FA-based approach are very large and much broader than those obtained with the FGD–based approach. For instance, at some dates – especially after periods of suddenly higher interest rate volatility – the length of the intervals provided by the FA–methodology is almost 200 basis points (see for instance the intervals around March 1, 2001 and March 1, 2002). Such confidence intervals lengths are too large for applied short-term risk management purposes. Moreover, they are also too conservative. Indeed, in a formal back-testing analysis not reported here the realized number of exceedances of confidence intervals produced by the FA–based approach was most of the times significantly lower than the one expected under the given confidence level. Intuitively, this conservative behavior happens because the part of volatility dynamics that is not filtered by the FA-based approach inflates the variability of the filtered interest rates residuals in the resulting historical simulation procedure: when bootstrapping such residuals to compute out-of-sample interest rate confidence intervals their excess variability generates too conservative interval estimates. Such effects arise also when applying FA to lower dimensional subsets of the maturities available in our data set. Since the interest rate interval estimates and the back-testing results implied by the FA-based approach are so poor, we do not discuss them in more detail in the rest of the paper. The bad short-term forecasting power of FA in combination with filtered historical simulation has not to be interpreted as a generic disdain for factor models. In fact, such models are well-known to be very useful, for instance for longer horizon yield curve prediction using macroeconomic based factors.

To end this section, we investigate whether the assumption of i.i.d. residuals needed in the historical simulation is fulfilled. When performing Ljung-Box tests for autocorrelations in the level and squares of residuals filtered using our FGD methodology up to 20th order for several different time-windows, we found in most cases no significance at the 5% confidence level and no significance overall at the 1% confidence level.

24

2.2 Yield curve confidence envelopes: some preliminary evidence

We examine and compare the out-of-sample performance and the accuracy of ahead confidence bounds for the yield curve, computed by means of three historical simulation-based procedures: the industry standard benchmark used by RiskMetrics$^{\text{TM}}$, one based on a standard multivariate AR-CCC-GARCH model dynamics and, finally, one based on the FGD approach. For any available time to maturity and any time in the back-testing sample we compute by historical simulation confidence intervals on the value of the corresponding future interest rates. By plotting these confidence bounds as a function of time to maturity we can obtain for each methodology a set of out-of-sample confidence "envelopes" for the whole yield curve at any relevant date. Examples of such yield curve confidence envelopes are presented in Figure 3, where we plot the realized yield curves at some given dates, together with the 95%-confidence envelopes obtained by means of filtered historical simulation based on the RiskMetrics$^{\text{TM}}$ approach (the dotted lines in Figure 3) and the FGD technique (the dash-dotted lines in Figure 3), respectively.

FIGURE 3 ABOUT HERE.

The term structure realizations presented in Figure 3 suggest at first sight that both methodologies yield reasonable confidence envelopes. In particular, in almost all graphs of Figure 3, the realized yield curves lie inside the corresponding 95%-confidence envelopes. A small exceedance of the FGD-based envelope bounds is observed for instance in the term structure on March 13, 2001, at weekly maturities. For the RiskMetrics$^{\text{TM}}$ approach one relatively large exceedance is observed on January 5, 2001, at the two months maturity. The FGD-based procedure seems to replicate better some particular shapes of the observed yield curves, especially at the shorter times to maturity. In some cases the term structure envelopes based on the RiskMetrics$^{\text{TM}}$ methodology appear to be too smooth as a function of time to maturity (see again for instance the graph in Figure 3 for the term structure on January 5, 2001).

In contrast to the results for the factor analysis, neither the RiskMetrics$^{\text{TM}}$ nor the CCC–based confidence intervals seem to be systematically more or less conservative than

25

those under the FGD-methodology from an unconditional perspective. To illustrate this point Figure 4 presents Box-Plots of the confidence interval lengths produced by RiskMetrics$^{\text{TM}}$ and by FGD for the ten years maturity interest rates and at confidence levels 95% and 99%.

FIGURE 4 ABOUT HERE.

From these graphs the median interval length under the RiskMetrics$^{\text{TM}}$ methodology appears to be lower than under FGD at the 95% and the 99% confidence levels. At the same time, the variability of the arising interval lengths for the RiskMetrics$^{\text{TM}}$ methodology is higher than for FGD, especially at the 95% confidence level. Such patterns are a direct consequence of the different implicit dynamic structures of confidence intervals estimated by means of the RiskMetrics$^{\text{TM}}$ and the FGD approaches. A comparison related to the one in Figure 4 is presented in Figure 5, presenting Box-Plots of the one-day up to 10 days ahead 95% confidence interval lengths estimated for the 5 years maturity interest rates by an AR-CCC-GARCH–based and a FGD-based methodology.

FIGURE 5 ABOUT HERE.

In Figure 5 we observe that the median interval lengths implied by an AR-CCC-GARCH approach are all lower than those implied by FGD. In contrast to the comparison with RiskMetrics$^{\text{TM}}$ in Figure 4, the AR-CCC-GARCH–based methodology implies also lower interquartile ranges of interval lengths than FGD for the 5 years maturity interest rate under scrutiny. However, the AR-CCC-GARCH–based approach also implies quite a few extreme interval lengths that are much larger than the corresponding ones under FGD (see for instance the top left panel in Figure 5). This is mainly a consequence of the higher time instability of interval lengths computed by the AR-CCC-GARCH–based methodology when large changes in interest rates occur; see also Section 2.5 below.

2.3   Back-testing one-day ahead confidence bounds

To compare more consistently and more precisely the effective performance or the above VaR prediction methodologies it is necessary to perform some more formal statistical

back-tests. To test the predictive performance of confidence envelopes of the yield curve we use two types of statistical tests, which are based on the frequency and the duration of yield curve envelope exceedances, i.e. the actual interest rate observations r_{t,T_i} that happen to fall outside the predicted confidence envelopes.

The first type of tests are standard overall frequency tests. Such tests test the hypothesis that the expected number of exceedances is compatible with the given confidence interval. For example, for a 95%-confidence envelope and a sample of 1000 back-testing days, one should expect 50 exceedances at any give time to maturity. In Table 2 we report for all methodologies under scrutiny the observed number of exceedances of one-day ahead confidence bounds for each time to maturity T_i, from 1 month to 30 years, i.e. $i = 4, .., 12$. For shorter times to maturity no methodology could provide accurate VaR estimation procedures in our sample. We report the observed number of exceedances at the confidence levels $0.9, 0.95, 0.99$ for the FGD-based methodology (CCC-FGD), the RiskMetrics[TM] approach (RM) and the historical simulation methodology based on a standard multivariate AR-CCC-GARCH dynamics (CCC). Under the null hypothesis, the observed number of exceedances is binomially distributed with a standard deviation ranging from 8.027 (for the 90%-confidence level) to 2.662 (for the 99%-confidence level). Back-testing results marked by one and two asterisks, respectively, denote a significant difference from the expected number of exceedances under the null hypothesis at the 5% and the 1% test nominal level, respectively.

TABLE 2 ABOUT HERE.

From Table 2, we observe that the FGD-based historical simulation strategy is the one that produces the lowest number of null hypothesis rejections when using overall frequency tests. In particular, for the 95% and the 99%-confidence envelopes we remark that only in one case a significant difference from the expected number of exceedances is observed. The RiskMetrics[TM] approach yields very often confidence intervals that are too tight and are therefore often violated a significantly larger number of times than expected under the null hypothesis. Similarly, also a standard CCC-GARCH-based historical simulation produces often too tight confidence intervals, especially for short and intermediate time

to maturities. Based on the results of pure overall frequency tests we conclude that the joint non-linear dependence of the yield curve dynamics estimated by FGD improves the accuracy of one day ahead interest rate confidence intervals computed by historical simulation.

A second type of test that can be applied in our back-testing exercise is a likelihood-ratio Weibull duration tests; see Christoffersen and Pelletier (2004). The basic idea of these tests relies on the fact that if a model for constructing the VaR confidence intervals at a confidence level q is correctly specified, then the conditional expected duration between consecutive exceedances - i.e. the expected no-hit duration - is constant and equal to $1/q$ days. Such an hypothesis can be tested as follows[12]. Let $D_j = t_j - t_{j-i}$ be the no-hit duration for time t_j, where t_j denotes the day of exceedance number j. Then, under the null hypothesis that the model is correctly specified, $E(D_j) = 1/q$ days for any $j = 1, 2, ...$ This hypothesis can be tested together with the independence hypothesis on the process of no-hit durations against some specific dependence alternative. To this end, we consider alternatives where the distribution of no-hit durations is a Weibull distribution with density given by

$$f_W(D; a, b) = a^b b D^{b-1} \exp\left(-(aD)^b\right),$$

where $a, b > 0$. The exponential distribution with parameter a then implies the only memoryless (continuous) random distribution in this class, which emerges as the special case $b = 1$. Thus, the null hypothesis of the likelihood-ratio Weibull duration test is

$$H_0 : \ b = 1 \text{ and } a = q, \tag{2.1}$$

where $b = 1$ is implied under the null hypothesis of independence. Let $\{C_j : \ j = 1, \dots, n\}$ be the hit sequence of $\{0, 1\}$ random variables that indicate if a no-hit duration D_j is censored ($C_j = 0$) or if it is not ($C_j = 1$).[13] For a given hit sequence and a given sequence of no-hit durations $D = \{D_j : j = 1, .., n\}$ the log-likelihood is given by

$$\log L(D; \theta) = (1 - C_1) \log\left(S(D_1)\right) + (1 - C_n) \log\left(S(D_n)\right) + \sum_{j=1}^{n} \left(C_j \log\left(f_W(D_j)\right)\right), \tag{2.2}$$

where in the case of a censored observation we merely know that no hit has been observed between time 0 and D_1 or between time $\sum_{j=1}^{n-1} D_j$ and D_n, respectively. In this case, the

contribution to the likelihood is given by the survival function $S(D_j) = \exp\left(-(aD_j)^b\right)$. The standard likelihood-ratio test statistic for testing (2.1) is then given by

$$\text{LR} = -2\left(\log L(D; \hat{a}, \hat{b}) - \log L(D; q, 1)\right), \tag{2.3}$$

where $\hat{a}, \hat{b}$ are the maximum likelihood estimators of the parameters a, b. This statistic is asymptotically chi-square distributed with two degrees of freedom.[14]

Results of the above likelihood-ratio Weibull duration tests for 1-day ahead yield curve confidence bounds are reported in Table 3 for our FGD-based historical simulation procedure (CCC-FGD), for the RiskMetrics[TM] one (RM) and for a multivariate AR-CCC-GARCH model based approach (CCC).

TABLE 3 ABOUT HERE.

As for the overall frequency tests an FGD-based historical simulation procedure is the one that clearly produces the lowest number of rejections of the relevant null hypothesis. Indeed, the only rejections are observed at the 95% confidence level for the one month and the six months times to maturity. The RiskMetrics[TM] approach yields confidence bounds which, especially for the 99% confidence level, are inconsistent with the hypothesis of independent durations between consecutive exceedances. The AR-CCC-GARCH model based approach produces 8 null hypothesis rejections at the different confidence levels, especially for time to maturities up to one year. These findings confirm that the joint non-linear dependence of the yield curve dynamics estimated by FGD improves the accuracy of VaR confidence intervals computed by historical simulation.

2.4 Back-testing confidence bounds for longer forecasting horizons

Accuracy of the above interest rates prediction methodologies for forecasting horizons longer than one day is investigated next. In this context, we found that for times to maturity up to about one year all historical simulation approaches under scrutiny produced a poor predictive power and inaccurate confidence interval estimates, with confidence

bounds that were often violated several times in a row. A more detailed data inspection showed that this is due principally to a sequence of multiple big interest rate shocks on the Euro market (often with changes larger than 0.3%-0.4%) caused by several adjustments in the Fed's target rate during the second part of our back-testing period. In the sequel we therefore focus on several days ahead interest rate predictions for longer terms to maturity between two years and thirty years. We remark, however, that interest rates in the short maturity spectrum still affect the forecasts of longer term interest rates, because they typically influence the conditional mean vector and the conditional covariance matrix in our estimated multivariate model for interest rate changes.

Results of overall frequency tests on the total number of exceedances at prediction horizons of 3,5 and 10 days are summarized in Table 4 for the FGD-based approach (CCC-FGD), the RiskMetrics^{TM} approach (RM) and the approach based on a multivariate AR-CCC-GARCH model (CCC).

TABLE 4 ABOUT HERE.

To correct for the autocorrelation in the series of exceedances under overlapping measurement intervals, we estimated the relevant standard errors using a Newey and West (1987) covariance matrix estimator with truncation parameter $x - 1$, where x is the forecasting horizon.

From Table 4 we see that also for longer forecasting horizons the FGD-based approach produces clearly better back-testing results, with only one null hypothesis rejection at the ten days forecasting horizon for the two years maturity interest rate. At the same time, the Riskmetrics^{TM} and the AR-CCC-GARCH methodologies provide a very bad back-testing performance, with 17 and 20 null hypothesis rejections, respectively, across the different forecasting horizons and confidence levels. These findings suggest that the joint non-linear dependence of the yield curve dynamics estimated by FGD improves even more crucially the VaR confidence intervals computed by historical simulation for longer forecasting horizons. Indeed, in terms of the pure number of null hypothesis rejections a standard AR-CCC-GARCH-based approach without FGD does not perform better in our study than a very simple Riskmetrics^{TM} approach. It is interesting to remark that the nonparametric

30

conditional mean and variance functions estimated by FGD for maturities from 2 to 30 years typically contain also lagged interest rates in the short-term spectrum of the yield curve. Therefore, the inclusion of such lagged short-term interest rates as instruments in a nonparametric FGD–approach enhances the quality of several days ahead interval predictions for longer term interest rates. A comparable quality in the forecasting ability of longer term interest rates could not be attained by means of (i) a three factor analysis of the yield curve dynamics, (ii) a Riskmetrics$^{\text{TM}}$–type approach or (iii) a parametric AR-CCC-GARCH–based historical simulation procedure.

We conclude the section by discussing how our methodology will perform for longer forecasting horizons. We performed the above analysis also for two-to-eight-weeks forecasting horizons and found that for horizons longer than four-five weeks all historical simulation approaches under scrutiny produced inaccurate confidence interval estimates. Therefore, the FGD historical simulation procedure based on daily data presented above is applicable for short term yield curve forecasting over horizons of about three weeks. This is not a surprising results, especially also because the stability of the bootstrap procedure in the historical simulation method can worsen dramatically as the horizon increases. A simple way of obtaining historical simulation scenarios over longer horizons would be, e.g., to estimate the model on weekly-data and to modify the required bootstrap procedure accordingly.

2.5   Confidence intervals for bond returns

An accurate yield curve scenarios generation technique can be used to compute the implied confidence intervals and risk measures for, in principle, any interest rate derivative. To this end, any generated yield curve scenario can be mapped into a corresponding derivative price scenario, to obtain an historically simulated distribution of derivative prices. For instance, such an historically simulated distribution of derivative prices can be used to compute the VaR or the Expected Shortfall of a derivative return for different short-term forecasting horizons. Such an exercise gives insight into the potential losses that are associated with the underlying interest rate risk factors. For derivative prices that depend on several interest rate points on the yield curve (for instance, a simple spread

portfolio) it is important to have a procedure generating accurate interest rates scenarios at the same time for (i) several interest rate maturities and (ii) several quantiles of the historical interest rate distribution.[15] From the empirical results in the last sections, the FGD–based historical simulation approach is the one which, among the methodologies studied in the paper, better fulfills these two requirements.

To illustrate the computation of the loss distribution for a simple derivative in the above historical simulation setting, consider the problem of computing three days ahead confidence intervals for the returns of a simple 10 years maturity US Treasury Notes. From the simulated 10 years interest rates we can easily compute the corresponding simulated three days bond returns and, from their simulated distribution, compute the associate return confidence intervals. The resulting dynamic three days ahead 99%-confidence intervals and the associated realized returns are presented in Figure 6 for (i) the Riskmetrics[TM]–type approach, (ii) the parametric AR-CCC-GARCH–based historical simulation procedure and (iii) the technique based on FGD.

Figure 6 ABOUT HERE.

In the top panel of Figure 6 the comparison between the Riskmetrics[TM] and the FGD-methodology shows that the first one tends to produce too liberal confidence intervals over time, especially in the lowest estimated quantiles of bond returns. This pattern causes a large number of realized losses that violate strongly and too often the estimated confidence bounds (see for instance the time periods around January 2001, May 2001 and November 2001). The realized number of exceedances (19) is consistent with the results in Table 4 and is significantly too high. The FGD–based methodology, instead, produces on average wider confidence bounds and less severe exceedances. The realized number of exceedances (7) is consistent with the results in Table 4 and is not statistically significantly different from the one expected under the null of a correct VaR prediction model.

The comparison between the AR-CCC-GARCH– and the FGD–based approaches is presented in the bottom panel of Figure 6. As a general remark, we observe an approximate tendency of both approaches to estimate confidence intervals with similar average interval lengths. However, the confidence intervals estimated by the AR-CCC-GARCH–based approach can happen to be too tight and are also more variable over time, especially

32

in periods of very variable interest rates and bond returns (see for instance the time period between May 2001 and May 2002). Such a higher variability of the estimated confidence interval lengths implies in some cases a too liberal confidence interval estimate and a corresponding back-testing exceedance. The realized number of exceedances (14) is consistent with the results in Table 4 and is statistically significantly different from the one expected under the null of a correct VaR prediction model.

3   Conclusions

We proposed a multivariate nonparametric technique based on FGD and historical simulation to generate more reliable scenarios and confidence intervals for the term structure of interest rates from historical data. The methodology is computationally feasible in large dimensions and can account for a non-linear time series dependence of interest rate at all available maturities. We back-tested our methodology on daily USD bond data and found that its out-of-sample accuracy is higher than the one of further scenario generating technologies based on factor analysis with conditional heteroskedastic factors, a multivariate AR-CCC-GARCH model, or the exponential smoothing covariance forecasting technique used by the RiskMetrics[TM] approach. At forecasting horizons of one day, FGD provided accurate multivariate VaR computations for time to maturities between one month and thirty years. For longer horizons (i.e. ten days) accurate VaR predictions are obtained for time to maturities between roughly one and thirty years.

Notes

[1] The incorporation of a possibly high number of further risk factors can be easily accomplished by FGD.

[2] In contrast to linear time series analysis, this assumption is a very standard one for non-linear models of the yield curve dynamics. It is necessary, because the yield levels enter in our model as predictor variables for the conditional mean and conditional covariance matrix of interest rate changes.

[3] If needed, the same procedure might also be applied for the conditional means, for example, by modeling the starting conditional mean estimates using an ARMA-type dynamics.

[4] Note that this kind of correlation targeting strategy to reduce the number of parameters to be estimated can not be used in a time-varying conditional correlation setting. By contrast, if the conditional correlation matrix varies through time, starting estimates for R_t can be constructed using standard two-stage procedures. Then, analogously to the CCC case, new estimates for R_t can be constructed iteratively by using the correlation dynamics specified in the different models. The computational costs, however, may not be neglected.

[5] Moreover, such higher lagged predictors enter only few times in the explicit construction of the additive term functions $g_{m,i}(\cdot)$ and $f_{m,i}(\cdot)$ based on a regression tree least squares fit.

[6] This line search guarantees that the negative log-likelihood is monotonically decreasing in the number of iteration steps.

[7] See Bollerslev (1990) for more details.

[8] This cross-validation scheme has been shown to work well in empirical applications of FGD; see again Audrino and Barone-Adesi (2005) and Audrino and Bühlmann (2003).

[9] In this sense, the estimated model could presumably imply some small arbitrage op-

portunities, which should be consistent, however, with the type of historical no-arbitrage violations that might be present in rough (non-smoothed) interest rate data.

[10]RiskMetrics[TM] uses an EWMA conditional variance estimator of the form

$$V_t = (1 - \lambda)\xi_{t-1}\xi_{t-1}^T + \lambda V_{t-1}, \ \lambda = 0.94, \quad (3.4)$$

where V_0 can be fixed to be the sample covariance matrix or some presample data selection used to initialize the smoother. This model is extremely easy to estimate since it contains only one parameter of interest. One obvious drawback is that it forces all assets to have the same smoothing coefficient $\lambda = 0.94$, irrespectively of the specific dynamic features of a given interest rate.

[11]Repeating the test for other windows of data we got similar results.

[12]See also Kiefer, 1988 or Gourieroux, 2000 for a general introduction to duration modeling.

[13]If the hit sequence $\{C_j \ j = 1, \ldots, n\}$ starts (ends) with 0 then D_1 (D_n) is the number of days until we get the first exceedance (number of days after the last exceedance) and $C_1 = 0$ $(C_n = 0)$. If instead the hit sequence starts (ends) with a 1, then $C_1 = 1$ and D_1 is simply the number of days until the second exceedance (then $C_n = 1$ and $D_n = t_n - t_{n-1}$).

[14]It is also possible to compute finite sample critical values for the above statistics by means of Monte Carlo simulation. Our results do not change in an essential way when doing that. We therefore further use standard asymptotic critical values.

[15]Clearly, not all interest rate derivatives will have a distribution of prices where upper and lower quantiles are associated only with the upper and lower quantiles of the interest rates affecting the derivative prices.

References

Alexander, C. (2001). A primer on the orthogonal GARCH model. ISMA Center, Mimeo.

Ang, A. and Piazzesi, M. (2003). A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. *Journal of Monetary Economics* **50**, Issue 4, 745-787.

Audrino, F. and Barone-Adesi, G. (2005). A multivariate FGD technique to improve VaR computation in equity markets. *Computational Management Science* **2**, 87-106.

Audrino, F. and Barone-Adesi, G. (2005). Functional gradient descent for financial time series with an application to the measurement of market risk. *Journal of Banking and Finance* **29**, 959-977.

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.

Barone-Adesi, G., Bourgoin, F. and Giannopoulos, K. (1998). Don't look back. *Risk* **11** (August), 100-104.

Barone-Adesi, G., Giannopoulos, K. and Vosper, L. (1999). VaR without correlations for portfolio of derivative securities. *Journal of Futures Markets* **19** (April), 583-602.

Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model. *The Review of Economics and Statistics* **72**, 498-505.

Bühlmann, P. (2006). Boosting for high-dimensional linear models. *Annals of Statistics* **34**, 559-583.

Bühlmann, P. (2003). Boosting methods: why they can be useful for high-dimensional data. Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003), Vienna, Austria.

Bühlmann, P., and Yu, B. (2003). Boosting with the L2 loss: regression and classification. *Journal of the American Statistical Association* **98**, 324-339.

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

Christoffersen, P. and Pelletier, D. (2004). Backtesting VaR: a duration based approach. *Journal of Financial Econometrics*, **2**, 84-108.

Diebold, F.X. and Li, C. (2006). Forecasting the term structure of government bond yields. *Journal of Econometrics* **130**, 337-364.

Diebold, F.X., Piazzesi, M. and Rudebusch, G. (2005). Modeling bond yields in finance and macroeconomics. *American Economic Review* **95**, 415-420.

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

Engle, R. F., Ng, V. K. and Rothschild, M. (1990). Asset pricing with a factor ARCH covariance structure: empirical estimates from treasury bills. *Journal of Econometrics* **54**, 2075-2107.

Freund, Y. and Schapire, R.E. (1996). Experiments with a new boosting algorithm. In Machine Learning: Proc. Thirteenth International Conference, pp. 148–156. Morgan Kauffman, San Francisco.

Friedman, J.H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of Statistics* **29**, 1189-1232.

Friedman, J.H., Hastie, T. and Tibshirani, R. (2000). Additive logistic regression: a statistical view of boosting. *Annals of Statistics* **28**, 337-407 (with discussion).

Gallant, A.R. and Tauchen, G. (1989). Seminonparametric estimation of conditionally constrained heterogeneous processes: asset pricing applications. *Econometrica* **57**, 1091-1120.

Gourieroux, C. (2000). *Econometrics of qualitative dependent variables.* Cambridge University Press.

Engle, R.F. and Sheppard, N. (2001). Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH. Working Paper, University of California, San Diego.

Jamshidian, F. and Zhu, Y. (1997). Scenario simulation model: theory and methodology. *Finance and Stochastics* **1**, 43-67.

Kiefer, N. (1988). Economic duration data and hazard functions. *Journal of Economic Literature* **26**, 646-679.

Lekkos, I. (2000). A critique of factor analysis of interest rates. *The Journal of Derivatives* **8**, 72-83.

Loretan, M. (1997). Generating market risk scenarios using principal component analysis: methodological and practical considerations. Manuscript, Federal Reserve Board.

Lugosi, G. and Vayatis, N. (2004). On the bayes-risk consistency of regularized boosting methods. *The Annals of Statistics* **31**, No. 1, 30-55.

Mancini, L., Ronchetti, E. and Trojani F. (2005). Optimal Conditionally Unbiased Bounded-Influence Inference in Dynamic Location and Scale Models. *Journal of the American Statistical Association* **100**, 628-641.

Mannor, S., Meir, R. and Zhang, T. (2003). Greedy algorithms for classification – consistency, convergence rates, and adaptivity. *Journal of Machine Learning Research* **4**, 713-741.

Mardia, K. V. (1979). *Multivariate Analysis*. Academic Press, London.

Reimers, M. and Zerbs, M. (1999). A multi-factor statistical model for interest rates. *Algo Research Quarterly*, Vol. 2, No. 3, 55-63.

Rodrigues, A. P. (1997). Term structure and volatility shocks. Manuscript, Federal Reserve Board.

Tse, Y.K. (2000). A test for constant correlations in a multivariate GARCH model. *Journal of Econometrics* **98**, 107-127.

Zhang, T. (2004). Statistical behavior and consistency of classification methods based on convex risk minimization. *The Annals of Statistics* **31**, No. 1, 56-134.

Zhang, T. and Yu, B. (2005). Boosting with early stopping: convergence and consistency. *Annals of Statistics* **33**, 15381579.
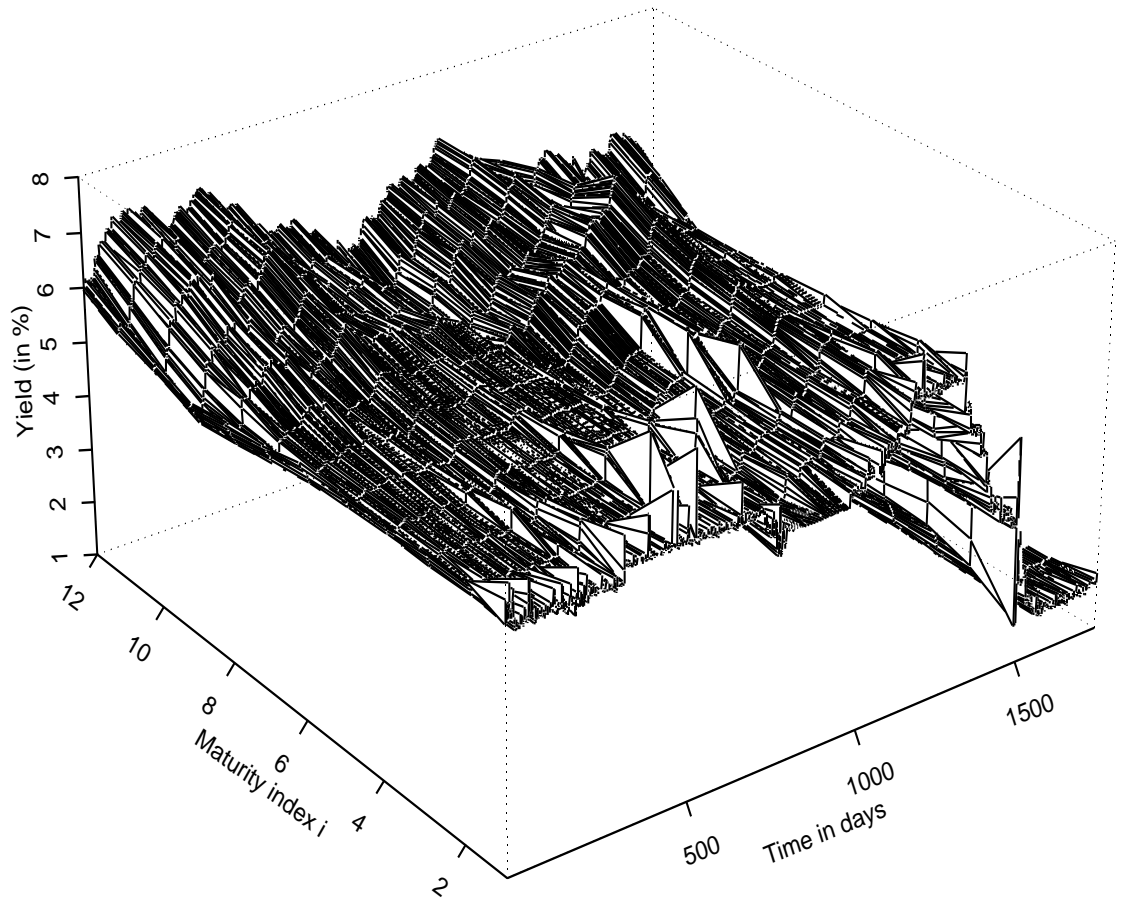
Figure 1: Term structure data: the sample consists of 1760 daily observations between January 1, 1996 and September 30, 2002 for twelve times to maturity $T_i=$ overnight, 1 week, 2 weeks, 1 month, 2 months, 3 months, 6 months, 1 year, 2 years, 5 years, 10 years, 30 years.

CCC-FGD vs. Factor Analysis
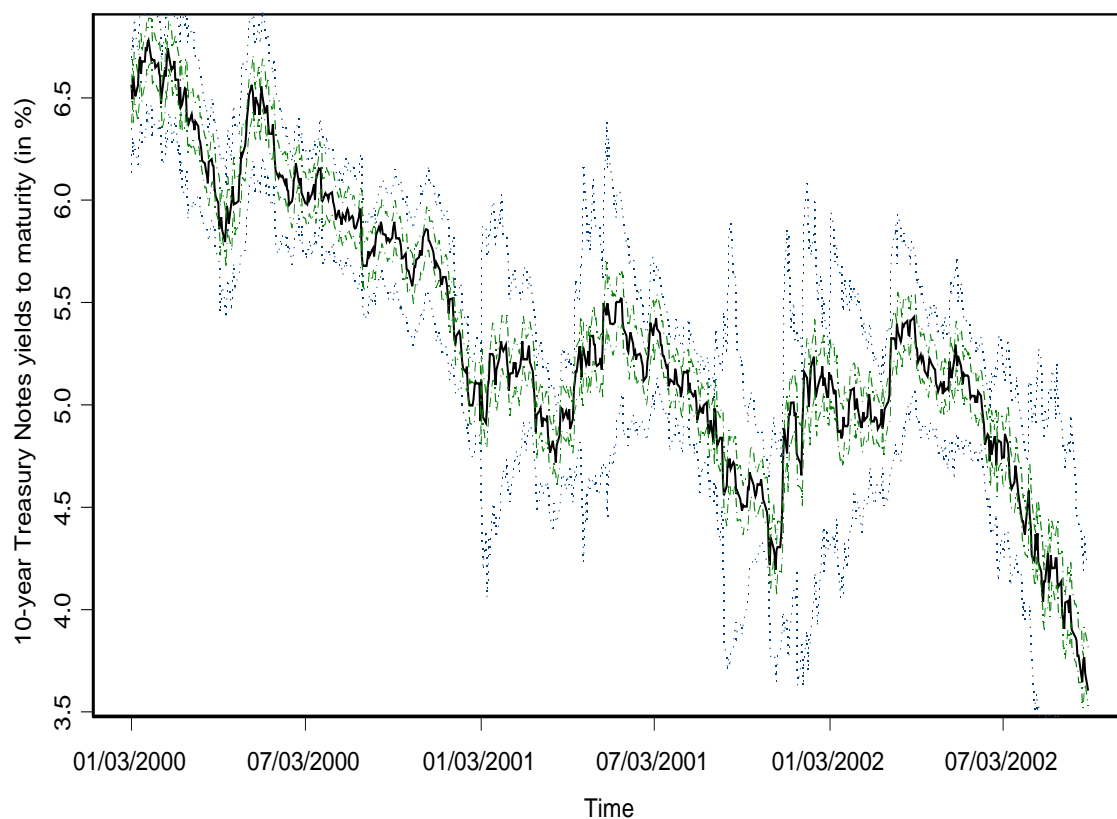
Figure 2: Out-of-sample one day ahead 95% interest rate interval estimates for the 10–year maturity bond. The straight line is the realized interest rate level. The two dotted lines are the estimated upper and lower interest rate quantiles when using the FA–based approach. The two dot dashed lines are the estimated upper and lower interest rate quantiles when using the FGD–based approach.

Maturity	sample mean	sample sdev	min	max	LB(20)	|LB(20)|
overnight	−0.0022	0.1374	−2.3200	1.5312	236.76*	450.44*
1 week	−0.0022	0.0701	−0.7501	1.3437	149.26*	291.54*
2 weeks	−0.0022	0.0606	−0.8331	0.8437	64.098*	177.50*
1 month	−0.0022	0.0724	−0.9800	1.0200	148.78*	260.16*
2 months	−0.0022	0.0422	−0.5800	0.9376	61.023*	179.52*
3 months	−0.0022	0.0354	−0.5900	0.6250	75.219*	302.35*
6 months	−0.0021	0.0382	−0.5500	0.2000	61.514*	318.80*
1 year	−0.0020	0.0567	−0.5312	0.6650	31.589*	189.59*
2 years	−0.0020	0.0607	−0.5190	0.3240	30.034	294.27*
5 years	−0.0016	0.0623	−0.3720	0.3400	29.634	200.26*
10 years	−0.0011	0.0592	−0.2240	0.3340	26.079	137.92*
30 years	−0.0007	0.0492	−0.3240	0.2460	20.816	54.186*

Table 1: Summary statistics on time series of interest rate changes (in %) at twelve different maturities for the time period between January 1, 1996 and September 30, 2002, for a total of 1760 observations. Sample sdev, LB(20) and |LB(20)| are the sample standard deviations and the Ljung-Box statistics testing for autocorrelation in the time series of interest rate changes and absolute interest rate changes, respectively, up to the 20th lag. Asterisks indicate statistical significance at the 5% confidence level.
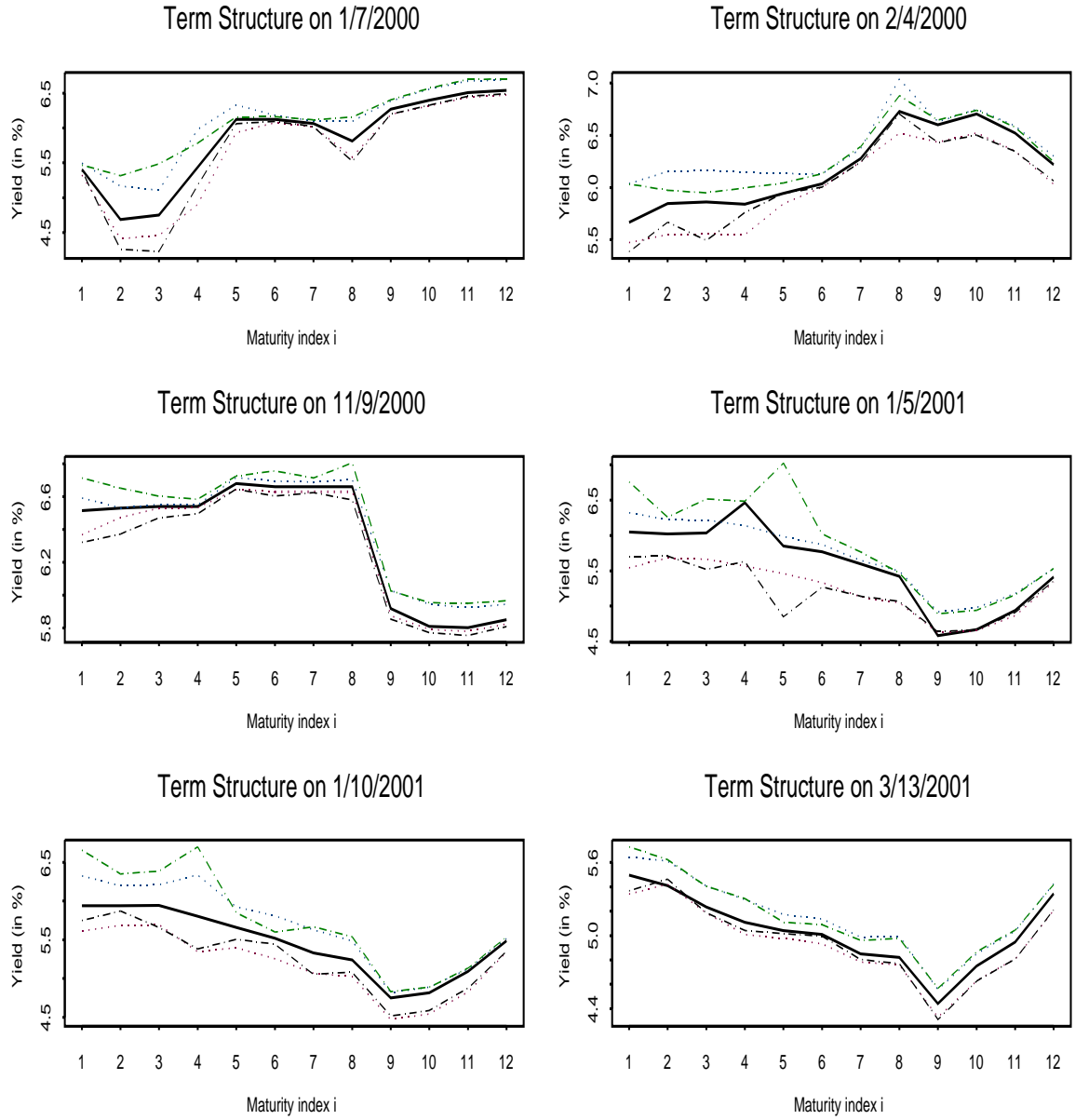
Figure 3: Realized yield curves (solid line) and one day ahead 95%-confidence envelopes using (i) the FGD–based setting (dot dashed lines) and (ii) the RiskMetrics^{TM} approach (dotted lines) in the estimation of conditional means and volatilities for the corresponding historical simulation procedure. The plotted yield curve envelopes are for some selected dates in the backtesting period from January 1, 2000, to September, 30, 2002. The maturity index $i = 1, .., 12$ in the graphs corresponds to twelve ordered maturities: overnight, 1 week, 2 weeks, 1 month, 2 months, 3 months, 6 months, 1 year, 2 years, 5 years, 10 years, 30 years.

43
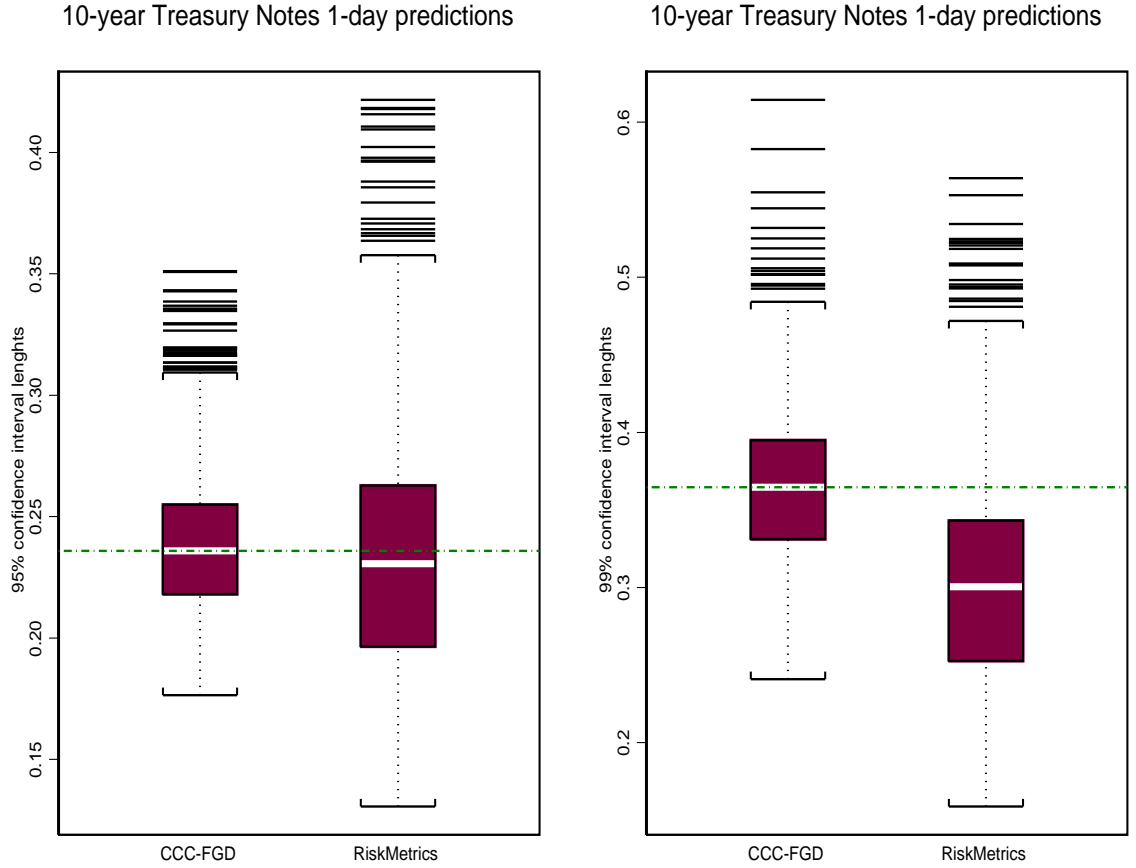
Figure 4: One day ahead confidence interval lengths for the 10–years maturity interest rate under a 95% (left panel) and a 99% (right panel) confidence level. In each panel, the right (left) Box Plot is for the RiskMetrics^{TM}–based (the FGD–based) approach.

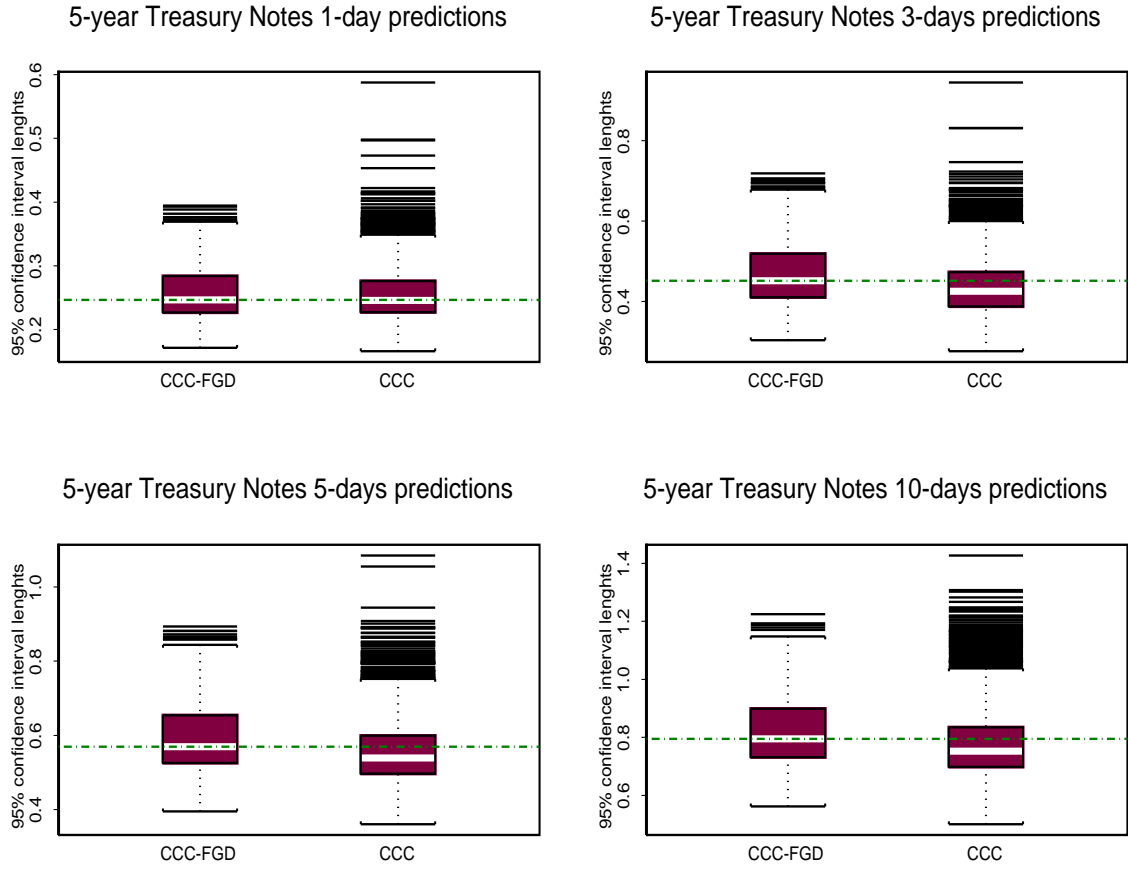Figure 5: One day (left top panel), three days (right top panel), five days (left bottom panel) and ten days (left right panel) ahead confidence interval lengths for the 5–years maturity interest rate under a 95% confidence level. In each panel, the right (left) Box Plot is for the CCC-AR-GARCH–based (the FGD–based) approach.
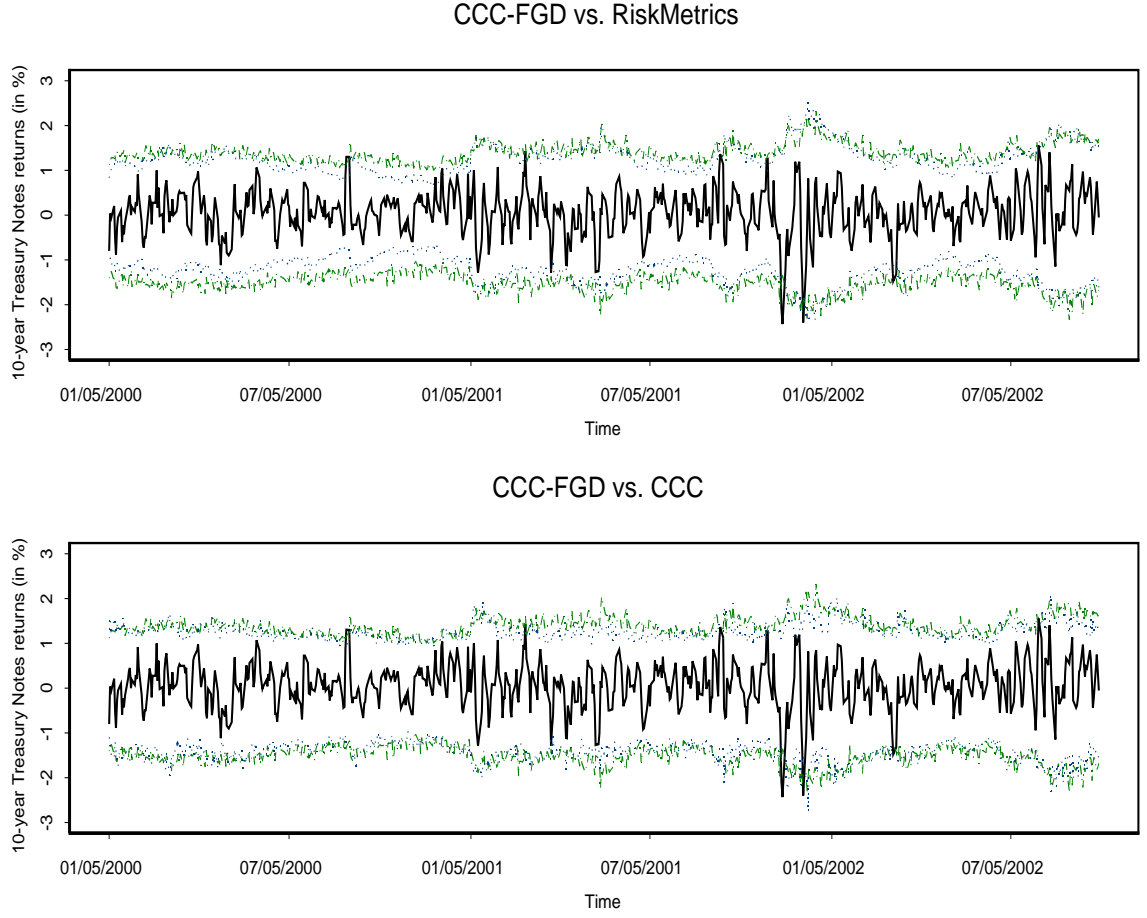
Figure 6: Three days ahead confidence intervals for the 10–years maturity zero bond return under a 99% confidence level. In each panel the straight line is the realized bond return. In the upper panel, the two dotted lines are the estimated upper and lower return quantiles when using the Riskmetrics^{TM}–based approach. The two dot dashed lines are the estimated upper and lower return quantiles when using the FGD–based approach. In the bottom panel, the two dotted lines are the estimated upper and lower return quantiles when using the parametric AR-CCC-GARCH–based approach. The two dot dashed lines in the bottom panel are again the estimated upper and lower return quantiles when using the FGD–based approach.

Maturity	Confidence level								
	90%			95%			99%		
	CCC-FGD	RM	CCC	CCC-FGD	RM	CCC	CCC-FGD	RM	CCC
Expected	71.6			35.8			7.16		
1 month	61	49**	86*	26	39	31	5	25**	3
2 months	75	59	99**	37	43	55**	8	22**	8
3 months	76	55*	104**	40	42	57**	7	26**	15**
6 months	77	74	88*	44	53**	47*	9	37**	13*
1 year	84	69	106**	53**	47*	56**	10	31**	13*
2 years	83	80	91*	39	50*	40	8	20**	4
5 years	87*	80	93**	39	49*	43	4	21**	6
10 years	91*	80	85	40	51**	43	7	22**	7
30 years	85	78	81	38	43	42	4	16**	6

Table 2: Overall frequency tests: exceedances for one-day ahead confidence bound forecasts recorded for times to maturity between one month and 30 years in the backtesting period from January 3, 2000 to September 30, 2002 (for a total of 716 trading days). The predictions are constructed using the FGD algorithm of Section 1 (CCC-FGD), the RiskMetrics$^{\text{TM}}$ approach (RM) and a standard multivariate AR-CCC-GARCH model. Results marked with one and two asterisks show significance at the 5% and the 1% confidence levels, respectively, for binomial tests investigating differences from the expected number of exceedances.

Maturity	Confidence level								
	90%			95%			99%		
	FGD	RM	CCC	FGD	RM	CCC	FGD	RM	CCC
1 month	5.58	13.9**	4.02	8.52*	8.06*	1.32	2.06	29.8**	5.82
2 months	0.08	6.99*	9.81**	0.17	2.95	9.13*	0.13	17.5**	0.02
3 months	5.78	5.01	12.4**	2.02	0.72	10.6**	0.30	27.2**	6.27*
6 months	0.47	1.85	3.31	1.36	7.30*	4.57	0.37	58.9**	2.91
1 year	7.05	0.19	13.8**	5.95	2.92	9.22**	0.61	40.9**	3.13
2 years	5.95	0.85	5.76	0.45	4.51	0.49	1.18	13.8**	3.14
5 years	3.12	1.13	6.14*	0.17	3.75	1.34	3.41	18.7**	1.39
10 years	4.46	0.80	2.77	0.57	5.06	1.31	1.85	18.4**	0.85
30 years	2.15	1.68	0.95	0.06	1.60	0.98	3.41	8.14*	1.36

Table 3: Likelihood-ratio Weibull duration tests: exceedances for one-day ahead confidence bound forecasts recorded for the same maturities of Table 2 in the backtesting period from January 3, 2000 to September 30, 2002 (for a total of 716 trading days). The predictions are constructed using the FGD algorithm of Section 1 (FGD), the RiskMetrics$^{\text{TM}}$ approach (RM) and a standard multivariate AR-CCC-GARCH model (CCC). Results marked with one and two asterisks show significance at the 5% and the 1% confidence levels, respectively.

3-day predictions

Maturity	Confidence level								
	90%			95%			99%		
	CCC-FGD	RM	CCC	CCC-FGD	RM	CCC	CCC-FGD	RM	CCC
Expected	71.6			35.8			7.16		
2 years	86	91*	92*	40	57**	56**	11	26**	14*
5 years	76	88	94*	41	62**	52*	11	23**	16*
10 years	69	75	84	30	42	53*	7	19**	14*
30 years	59	75	74	29	40	36	10	15*	9

5-day predictions

Maturity	Confidence level								
	90%			95%			99%		
	CCC-FGD	RM	CCC	CCC-FGD	RM	CCC	CCC-FGD	RM	CCC
2 years	82	96*	100*	43	58*	55*	16	27**	17*
5 years	80	96*	104**	43	64**	58*	16	26**	17*
10 years	65	81	87	32	39	48	7	15	9
30 years	64	80	78	42	44	50	8	13	11

10-day predictions

Maturity	Confidence level								
	90%			95%			99%		
	CCC-FGD	RM	CCC	CCC-FGD	RM	CCC	CCC-FGD	RM	CCC
2 years	86	99	108*	54	68*	65*	26*	35**	28*
5 years	81	97	112*	46	63*	66*	21	36**	26*
10 years	67	82	94	32	50	45	8	17	10
30 years	62	81	77	38	51	45	12	21	14

Table 4: Overall frequency tests: number of exceedances for 3-days (top panel), 5-days (middle panel) and 10-days (bottom panel) ahead confidence bound forecasts recorded for maturities between 2 and 30 years in the back-testing period from January 3, 2000 to September 30, 2002 (for a total of 716 trading days). The predictions are constructed using the FGD algorithm of Section 1 (CCC-FGD), the risk RiskMetrics[TM] approach (RM) and a standard multivariate AR-CCC-GARCH model (CCC). Results marked with one and two asterisks show significance at the 5% and the 1% confidence level, respectively, for binomial tests investigating differences from the expected number of exceedances. Standard errors have been computed by means of a Newey and West (1987) covariance matrix estimator to correct for the autocorrelation in the exceedances time series.