Available online at www.sciencedirect.com
ScienceDirect
International Journal of Forecasting 25 (2009) 304–327
www.elsevier.com/locate/ijforecast

Asymmetric effects and long memory in the volatility of Dow Jones stocks

Marcel Scharth[a], Marcelo C. Medeiros[b,*]

[a] Department of Economics, The University of Amsterdam, Netherlands
[b] Department of Economics, Pontifical Catholic University of Rio de Janeiro, Brazil

Abstract

Does volatility reflect a continuous reaction to past shocks or do changes in the markets induce shifts in the volatility dynamics? In this paper, we provide empirical evidence that cumulated price variations convey meaningful information about multiple regimes in the realized volatility of stocks, where large falls (rises) in prices are linked to persistent regimes of high (low) variance in stock returns. Incorporating past cumulated daily returns as an explanatory variable in a flexible and systematic nonlinear framework, we estimate that falls of different magnitudes over less than two months are associated with volatility levels 20% and 60% higher than the average of periods with stable or rising prices. We show that this effect accounts for large empirical values of long memory parameter estimates. Finally, we show that, while introducing more realistic dynamics for volatility, the model is able to overall improve or at least retain out-of-sample performance in forecasting when compared to standard methods. Most importantly, the model is more robust to periods of financial crises, when it attains significantly better forecasts.

Keywords: Realized volatility; Long memory; Asymmetric effects; Regime switching; Regression trees; Smooth transition; Forecasting; Empirical finance

1. Introduction

Does stock return volatility reflect a long-lived reaction to past shocks, or do structural breaks induce shifts in the volatility dynamics? Long range dependence (highly persistent autocorrelations) is a well documented stylized fact of the volatility of financial time series. This effect was first analyzed by Taylor (1986) for absolute values of stock returns. Ding, Granger, and Engle (1993) and de Lima and Crato (1993) considered powers of returns. More recently, Andersen, Bollerslev, Diebold, and Ebens (2001) studied the case of realized volatility.[1] Even

* Corresponding author.
E-mail addresses: marcelsfp@gmail.com (M. Scharth), mcm@econ.puc-rio.br (M.C. Medeiros).

[1] Realized variance is defined as the sum of squared intraday returns sampled at a sufficiently high frequency, consistently

0169-2070/$ - see front matter
doi:10.1016/j.ijforecast.2009.01.008

though the traditional GARCH (Generalized Autoregressive Conditional Heteroscedasticity) models of Engle (1982) and Bollerslev (1986) are able to describe the recurrent clusters in volatility, the short run dynamics of those models were shown to be an incomplete description of the data. Volatility breeds volatility; but could volatility today reflect a particularly volatile week a year ago? How do markets keep the memory of past movements? Furthermore, is there any explanation for the long-range dependence?

In this paper, we propose a novel approach to modeling and forecasting volatility by considering the possibility of structural changes and regime switches in volatility dynamics. We inquire how *changes in the markets* affect volatility. The goal is to link regime switches to long-range dependence, as well as to provide empirical evidence that long-term price variations convey meaningful information about multiple regimes in the realized volatility of stocks. From the asymmetric effects literature, it is known that negative returns are related to subsequent increases in volatility. Econometric models such as Nelson's (1991) Exponential GARCH (EGARCH) and the GJR-GARCH of Glosten, Jagannathan, and Runkle (1993) have been proposed to capture this effect. Nevertheless, the literature so far has focused almost exclusively on the relationship observed over one or a few days. For example, Andersen et al. (2001) ran a regression with a lagged negative return dummy, and concluded that the economic impact of the leverage effect on the realized variance of stocks belonging to the Dow Jones Industrial Average Index (DJIA) is marginal. An exception is Bollerslev, Litvinova, and Tauchen (2005), who examined evidence on the negative correlation between stock market movements and stock market volatility over intraday sampling frequencies.

Focusing on the realized volatility (RV) series of sixteen Dow Jones Industrial Average (DJIA) stocks over the period from 1994 to 2003, we consider the following questions: Are volatility levels the same in periods of significant losses, like the end of 2002 (when the DJIA reached a 4 year bottom), and periods of significant gains, like the year 2003 (when the DJIA went up 25%)? Can negative returns over some horizons be associated with regimes of higher volatility? We pursue the argument by incorporating past cumulated daily returns in the modeling framework of volatility series. Then, if price variations matter, what are the magnitudes that can be associated with regime switching behavior? And what are the relevant horizons? To tackle these considerations, our econometric strategy is developed around a flexible and systematic modeling cycle based on the tree-structured smooth transition regression model (STR-Tree) of da Rosa, Veiga, and Medeiros (2008). We choose a particular set of series in order to represent the most important components of the DJIA.

Our main result is that the effect on volatility of falls and rises in prices is in fact highly significant, and accounts for the evidence of long-range dependence in volatility, even in samples spanning several years. For example, we show that the daily volatility series of the IBM stock can be described by a nonlinear model where falls of various magnitudes over less than two months are associated with volatility levels 20% and 60% higher than the average of periods with stable or rising prices. Based on those findings, we propose a new model to describe and forecast realized volatility. When compared with alternative specifications with short and long memory, the more realistic model proposed in this paper is able to at least retain, and in some cases improve, the overall out-of-sample forecasting performance. Most importantly, the model is more robust to periods of financial crises and high volatility (which are the crucial ones from the point of view of risk management), when it attains significantly better forecasts. A model that allows for smoothly changing parameters across time (in order to capture possible structural breaks) is also estimated. However, the regime switching mechanism controlled by past cumulated returns turns out to be statistically superior. The results are uniform across 15 of the 16 series

approximating the integrated variance over the fixed interval where the observations are summed. Realized volatility is the squared root of the realized variance. In practice, high frequency measures are contaminated by microstructure noise such as bid-ask bounce, asynchronous trading, infrequent trading, and price discreteness, among others; see Biais, Glosten, and Spatt (2005). Ignoring the remaining measurement error, this ex post volatility measure can be modeled as an "observable" variable, in contrast to the latent variable models. See Andersen, Bollerslev, Diebold, and Labys (2003) and Barndorff-Nielsen and Shephard (2002) for the theoretical foundations of realized volatility. Several recent papers have proposed corrections to the estimation of RV in order to take the microstructure noise into account; see McAleer and Medeiros (2008) for a review. In this paper, we refer to realized volatility as a consistent estimator of the squared root of the integrated variance.

considered in this paper. Another important empirical finding is that, in terms of forecast ability, a simple exponentially weighted moving average model applied to the realized volatility series exhibits competitive behavior when longer horizons are considered. This is in part explained by the low and persistently decaying volatility at the end of the sample.

The rest of the paper is structured as follows. Section 2 briefly discusses the tree-structured smooth transition regression model, describing the inference procedures, model building strategy and estimation. In Section 3, we describe the data and the specification of our model, and present the estimations for models with structural breaks and asymmetric effects. The relationship between asymmetric effects and long memory is investigated in Section 3.2. Section 3.3 contains an analysis of point forecasting performances, and Section 4 concludes.

2. Modeling framework

In this section, we present the non-linear econometric model used in the paper. The discussion is partially based on da Rosa et al. (2008).

2.1. A brief introduction to regression trees

Let $\mathbf{x}_t = (x_{1t}, \ldots, x_{qt})' \in \mathbb{X} \subseteq \mathbb{R}^q$ be a vector which contains q explanatory variables (covariates or predictor variables) for a continuous univariate response $y_t \in \mathbb{R}$, $t = 1, \ldots, T$. Suppose that the relationship between y_t and $\mathbf{x}_t$ follows a regression model of the form

$$y_t = f(\mathbf{x}_t) + \varepsilon_t, \tag{1}$$

where the function $f(\cdot)$ is unknown, and, in principle, there are no assumptions about the distribution of the random term ε_t, apart from $\mathbb{E}(\varepsilon_t|\mathbf{x}_t) = 0$. A regression tree is a nonparametric model based on the recursive partitioning of the covariate space $\mathbb{X}$, which approximates the function $f(\cdot)$ as a sum of local models, each of which is determined in $K \in \mathbb{N}$ different regions (partitions) of $\mathbb{X}$. The model is usually displayed in a graph which has the format of a binary decision tree, with $N \in \mathbb{N}$ parent (or split) nodes and $K \in \mathbb{N}$ terminal nodes (also called leaves), and which grows from the root node to the terminal nodes. Usually, the partitions are defined by a set of hyperplanes, each of which is orthogonal to the axis of a given predictor variable, called the *split variable*. The most important reference in regression tree models is the Classification and Regression Trees (CART) approach put forward by Breiman, Friedman, Olshen, and Stone (1984). In this context, the local models are just constants.

To mathematically represent a regression-tree model, we introduce the following notation. The root node is at position 0, and a parent node at position j generates left- and right-child nodes at positions $2j+1$ and $2j + 2$, respectively. Every parent node has an associated split variable $x_{s_j t} \in \mathbf{x}_t$, where $s_j \in \mathbb{S} = \{1, 2, \ldots, q\}$. Furthermore, let $\mathbb{J}$ and $\mathbb{T}$ be the sets of indexes of the parent and terminal nodes, respectively. Then, a tree architecture can be fully determined by $\mathbb{J}$ and $\mathbb{T}$.

Example 1. Consider a regime switching volatility model that allows for multiple regimes associated with asymmetric effects, where the influence of a negative return on volatility for the next day depends on the behavior of returns over the past week. Define $r_{5,t}$ as the cumulated return over a horizon of five days and r_t as the daily return. Suppose that the daily volatility (σ_t) follows a piecewise constant process, where the conditional mean depends on the sign of the return on the previous day. This effect itself is weaker in "good weeks" (or a positive return over the last five days) than in "bad weeks" (or a negative return over the last five days), such that $\sigma_t = \omega_1 + \varepsilon_t$ if $r_{t-1} \geq 0$, $\sigma_t = \omega_2 + \varepsilon_t$ if $r_{t-1} < 0$ and $r_{5,t-1} \geq 0$ and $\sigma_t = \omega_3 + \varepsilon_t$ if $r_{t-1} < 0$ and $r_{5,t-1} < 0$. ε_t is white noise, and $\omega_3 > \omega_2 > \omega_1$ are constants. This model can be described in the regression tree with two parent nodes at positions 0 and 2 ($N = 2$, $\mathbb{J} = \{0, 2\}$) and three leaves or terminal nodes at positions 1,5 and 6 ($K = 3$, $\mathbb{T} = \{1, 5, 6\}$). See Fig. 1.

2.2. Tree-structured smooth transition regression

The STR-Tree model is an extension of the regression tree model, where the sharp splits are replaced by smooth splits given by a logistic function defined as

$$G(x; \gamma, c) = \frac{1}{1 + e^{-\gamma(x-c)}}. \tag{2}$$

The parameter γ, called the *slope parameter*, controls the smoothness of the logistic function. The regression

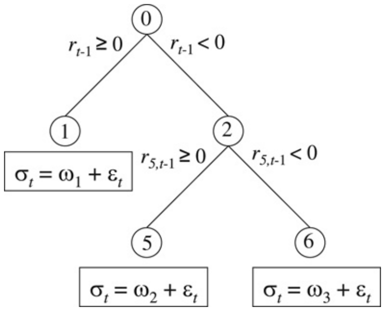

Fig. 1. Graphical representation of the volatility model described in Example 1.

tree model is nested in the smooth transition specification as a special case obtained when the slope parameter approaches infinity. The parameter c is called the *location parameter*.

Define $\log(RV_t)$ as the logarithm of the daily realized volatility. In this paper, $\log(RV_t)$ follows an augmented specification of the STR-Tree model defined as follows.

Definition 1. Let $\mathbf{z}_t \subseteq \mathbf{x}_t$, such that $\mathbf{x}_t$ is defined as in Eq. (1) and $\mathbf{z}_t \in \mathbb{R}^p$, $p \leq q$. The sequence of real-valued vectors $\{\mathbf{z}_t\}_{t=1}^T$ is stationary and ergodic. Set $\tilde{\mathbf{z}}_t = (1, \mathbf{z}_t)'$ and $\mathbf{w}_t \in \mathbb{R}^d$ is a vector of linear regressors, such that $\mathbf{w}_t \not\subseteq \mathbf{x}_t$. The time series $\{\log(RV_t)\}_{t=1}^T$ follows a Smooth Transition Regression Tree model, STR-Tree, if

$$\begin{aligned}\log(RV_t) &= H_{\mathbb{JT}}(\mathbf{x}_t, \mathbf{w}_t; \boldsymbol{\psi}) + \varepsilon_t \\ &= \boldsymbol{\alpha}'\mathbf{w}_t + \sum_{i \in \mathbb{T}} \boldsymbol{\beta}_i' \tilde{\mathbf{z}}_t B_{\mathbb{J}i}(\mathbf{x}_t; \boldsymbol{\theta}_i) + \varepsilon_t \end{aligned} \tag{3}$$

where

$$\begin{aligned} B_{\mathbb{J}i}(\mathbf{x}_t; \boldsymbol{\theta}_i) &= \prod_{j \in \mathbb{J}} G(x_{s_j,t}; \gamma_j, c_j)^{\frac{n_{i,j}(1+n_{i,j})}{2}} \\ &\times \left[1 - G(x_{s_j,t}; \gamma_j, c_j)\right]^{(1-n_{i,j})(1+n_{i,j})} \end{aligned} \tag{4}$$

and

$$n_{i,j} = \begin{cases} -1 & \text{if the path to leaf } i \text{ does not include the parent node } j; \\ 0 & \text{if the path to leaf } i \text{ includes the right-child node of the parent node } j; \\ 1 & \text{if the path to leaf } i \text{ includes the left-child node of the parent node } j, \end{cases} \tag{5}$$

where $H_{\mathbb{JT}}(\mathbf{x}_t, \mathbf{w}_t; \boldsymbol{\psi}) : \mathbb{R}^{q+1} \times \mathbb{R}^d \rightarrow \mathbb{R}$ is a nonlinear function indexed by the vector of parameters $\boldsymbol{\psi} \in \boldsymbol{\Psi}$ and $\{\varepsilon_t\}$ is a martingale difference sequence. Let $\mathbb{J}_i$ be the subset of $\mathbb{J}$ containing the indexes of the parent nodes that form the path to leaf i. Then, $\boldsymbol{\theta}_i$ is the vector containing all the parameters (γ_k, c_k) such that $k \in \mathbb{J}_i$, $i \in \mathbb{T}$.

The functions $B_{\mathbb{J}i}(\mathbf{x}_t; \boldsymbol{\theta}_i)$, $0 < B_{\mathbb{J}i}(\mathbf{x}_t; \boldsymbol{\theta}_i) < 1$, are known as *membership functions*, and it is easy to show that $\sum_{i \in \mathbb{T}} B_{\mathbb{J}i}\left(\mathbf{x}_t; \boldsymbol{\theta}_j\right) = 1$, $\forall\, \mathbf{x}_t \in \mathbb{R}^{q+1}$.

The parameters of Eq. (3) are estimated by nonlinear least-squares (NLS), which is equivalent to quasi-maximum likelihood estimation. Let $\hat{\boldsymbol{\psi}}$ be the quasi-maximum likelihood estimator (QMLE) of $\boldsymbol{\psi}$ given by

$$\begin{aligned}\hat{\boldsymbol{\psi}} &= \underset{\boldsymbol{\psi} \in \boldsymbol{\Psi}}{\operatorname{argmin}}\, \mathcal{Q}_T(\boldsymbol{\psi}) = \underset{\boldsymbol{\psi} \in \boldsymbol{\Psi}}{\operatorname{argmin}} \frac{1}{T}\sum_{t=1}^{T} q_t(\boldsymbol{\psi}) \\ &= \underset{\boldsymbol{\psi} \in \boldsymbol{\Psi}}{\operatorname{argmin}} \left\{ \frac{1}{T}\sum_{t=1}^{T} \left[\log(RV_t) \right.\right. \\ &\qquad \left.\left. - H_{\mathbb{JT}}(\mathbf{x}_t, \mathbf{w}_t; \boldsymbol{\psi})\right]^2 \right\}. \end{aligned} \tag{6}$$

Under stationarity of $\log(RV_t)$ and the identification of the STR-Tree model, it is straightforward to show that the estimator in Eq. (6) is consistent and asymptotically normal; see da Rosa et al. (2008) and Hillebrand and Medeiros (2008) for details.

2.3. Growing the tree

In this section we briefly present the modeling cycle adopted in this paper. The choice of relevant variables, the selection of the node to be split (if applicable), and the selection of the splitting (or transition) variable are carried out by a sequence of Lagrange Multiplier (LM) tests, following the ideas originally presented by Luukkonen, Saikkonen, and Teräsvirta (1988) and widely used in the literature.[2]

Consider that $\log(RV_t)$ follows a STR-Tree model with K leaves and we want to test whether the terminal

[2] See, for example, Teräsvirta (1994), van Dijk, Franses, and Paap (2002), van Dijk, Teräsvirta, and Franses (2002), or Medeiros and Veiga (2009).

node $i^* \in \mathbb{T}$ should be split or not. Write the model as

$$\begin{aligned}\log(RV_t) = \boldsymbol{\alpha}'\mathbf{w}_t + \sum_{i\in\mathbb{T}-\{i^*\}} \boldsymbol{\beta}_i'\tilde{\mathbf{z}}_t B_{\mathbb{J}i}(\mathbf{x}_t;\boldsymbol{\theta}_i) \\ + \boldsymbol{\beta}_{2i^*+1}'\tilde{\mathbf{z}}_t B_{\mathbb{J}2i^*+1}(\mathbf{x}_t;\boldsymbol{\theta}_{2i^*+1}) \\ + \boldsymbol{\beta}_{2i^*+2}'\tilde{\mathbf{z}}_t B_{\mathbb{J}2i^*+2}(\mathbf{x}_t;\boldsymbol{\theta}_{2i^*+2}) + \varepsilon_t,\end{aligned} \quad (7)$$

where

$$\begin{aligned}&B_{\mathbb{J}2i^*+1}(\mathbf{x}_t;\boldsymbol{\theta}_{2i^*+1}) = B_{\mathbb{J}i^*}(\mathbf{x}_t;\boldsymbol{\theta}_{i^*})\, G(x_{i^*t};\gamma_{i^*},c_{i^*})\\ &B_{\mathbb{J}2i^*+2}(\mathbf{x}_t;\boldsymbol{\theta}_{2i^*+2})\\ &\quad = B_{\mathbb{J}i^*}(\mathbf{x}_t;\boldsymbol{\theta}_{i^*})\,[1 - G(x_{i^*t};\gamma_{i^*},c_{i^*})].\end{aligned}$$

In a more compact form, Eq. (7) may be written as

$$\begin{aligned}\log(RV_t) = \boldsymbol{\alpha}'\mathbf{w}_t + \sum_{i\in\mathbb{T}-\{i^*\}} \boldsymbol{\beta}_i'\tilde{\mathbf{z}}_t B_{\mathbb{J}i}(\mathbf{x}_t;\boldsymbol{\theta}_i) \\ + \boldsymbol{\phi}_1'\tilde{\mathbf{z}}_t B_{\mathbb{J}i^*}(\mathbf{x}_t;\boldsymbol{\theta}_{i^*}) \\ + \boldsymbol{\phi}_2'\tilde{\mathbf{z}}_t B_{\mathbb{J}i^*}(\mathbf{x}_t;\boldsymbol{\theta}_{i^*})\, G(x_{i^*t};\gamma_{i^*},c_{i^*}) + \varepsilon_t,\end{aligned} \quad (8)$$

where $\boldsymbol{\phi}_1 = \boldsymbol{\beta}_{2i^*+2}$ and $\boldsymbol{\phi}_2 = \boldsymbol{\beta}_{2i^*+1} - \boldsymbol{\beta}_{2i^*+2}$.

In order to test the statistical significance of the split, a convenient null hypothesis is $\mathcal{H}_0 : \gamma_{i^*} = 0$ against the alternative $\mathcal{H}_a : \gamma_{i^*} > 0$. An alternative null hypothesis is $\mathcal{H}_0' : \phi_2 = 0$. However, it is clear from Eq. (8) that under $\mathcal{H}_0$, the nuisance parameters ϕ_2 and c_{i^*} can assume different values without changing the likelihood function, posing an identification problem; see Davies (1977, 1987).

A solution to this problem, proposed by Luukkonen et al. (1988), is to approximate the logistic function by a third-order Taylor expansion around $\gamma_{i^*} = 0$. After some algebra, we get

$$\begin{aligned}\log(RV_t) = \boldsymbol{\alpha}'\mathbf{w}_t + \sum_{i\in\mathbb{T}-\{i^*\}} \boldsymbol{\beta}_i'\tilde{\mathbf{z}}_t B_{\mathbb{J}i}(\mathbf{x}_t;\boldsymbol{\theta}_i) \\ + \boldsymbol{\alpha}_0'\tilde{\mathbf{z}}_t B_{\mathbb{J}i^*}(\mathbf{x}_t;\boldsymbol{\theta}_{i^*}) + \boldsymbol{\alpha}_1'\tilde{\mathbf{z}}_t B_{\mathbb{J}i^*}(\mathbf{x}_t;\boldsymbol{\theta}_{i^*})\, x_{i^*t} \\ + \boldsymbol{\alpha}_2'\tilde{\mathbf{z}}_t B_{\mathbb{J}i^*}(\mathbf{x}_t;\boldsymbol{\theta}_{i^*})\, x_{i^*t}^2 + \boldsymbol{\alpha}_3'\tilde{\mathbf{z}}_t B_{\mathbb{J}i^*}(\mathbf{x}_t;\boldsymbol{\theta}_{i^*})\, x_{i^*t}^3 \\ + e_t,\end{aligned} \quad (9)$$

where $e_t = \varepsilon_t + \phi_2 B_{\mathbb{J}i^*}(\mathbf{x}_t;\boldsymbol{\theta}_{i^*})\, R(x_{i^*t};\gamma_{i^*},c_{i^*})$ and $R(x_{i^*t};\gamma_{i^*},c_{i^*})$ is the remainder. The parameters $\boldsymbol{\alpha}_k$, $k = 0,\ldots,3$ are functions of the original parameters of the model.

Thus, the null hypothesis becomes

$$\mathcal{H}_0 : \boldsymbol{\alpha}_1 = \boldsymbol{\alpha}_2 = \boldsymbol{\alpha}_3 = 0. \quad (10)$$

Under $\mathcal{H}_0$, $R(x_{i^*t};\gamma_{i^*},c_{i^*}) = 0$ and $e_t = \varepsilon_t$, such that the properties of the error process remain unchanged under the null, and thus asymptotic inference can be used. The test statistic is given by[3]:

$$\begin{aligned}LM = \frac{1}{\hat{\sigma}^2}\sum_{t=1}^T \hat{u}_t\hat{\boldsymbol{\nu}}_t' \left\{\sum_{t=1}^T \hat{\boldsymbol{\nu}}_t\hat{\boldsymbol{\nu}}_t' - \sum_{t=1}^T \hat{\boldsymbol{\nu}}_t\hat{\mathbf{h}}_t' \right.\\ \left.\times\left(\sum_{t=1}^T \hat{\mathbf{h}}_t\hat{\mathbf{h}}_t'\right)^{-1} \sum_{t=1}^T \hat{\mathbf{h}}_t\hat{\boldsymbol{\nu}}_t'\right\}^{-1} \sum_{t=1}^T \hat{\boldsymbol{\nu}}_t\hat{u}_t,\end{aligned} \quad (11)$$

where $\hat{u}_t = y_t - H_{\mathbb{J}\mathbb{T}}(\mathbf{x}_t,\mathbf{w}_t;\hat{\boldsymbol{\psi}})$, $\hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^T \hat{u}_t^2$, $\hat{\mathbf{h}}_t = \frac{\partial H_{\mathbb{J}\mathbb{T}}(\mathbf{x}_t,\mathbf{w}_t;\boldsymbol{\psi})'}{\partial\boldsymbol{\psi}}|_{\mathcal{H}_0}$, and

$$\begin{aligned}\hat{\boldsymbol{\nu}}_t = \Big[&\tilde{\mathbf{z}}_t B_{\mathbb{J}i^*}\left(\mathbf{x}_t;\hat{\boldsymbol{\theta}}_{i^*}\right) x_{i^*t}, \tilde{\mathbf{z}}_t B_{\mathbb{J}i^*}\left(\mathbf{x}_t;\hat{\boldsymbol{\theta}}_{i^*}\right) x_{i^*t}^2,\\ &\tilde{\mathbf{z}}_t B_{\mathbb{J}i^*}\left(\mathbf{x}_t;\hat{\boldsymbol{\theta}}_{i^*}\right) x_{i^*t}^3\Big]'.\end{aligned}$$

Under $\mathcal{H}_0$, LM has an asymptotic χ^2 distribution with $m = 3(p+1)$ degrees of freedom.

As the assumption of normal and homoskedastic errors is usually violated in financial data, we carry out a robust version of the LM test, following the results of Wooldridge (1990). The test is implemented as follows:

(1) Estimate the model with K regimes. If the sample size is small and the model is thus difficult to estimate, numerical problems in applying the maximum likelihood algorithm may lead to a solution such that the residual vector is not precisely orthogonal to the gradient matrix of $H_{\mathbb{J}\mathbb{T}}(\mathbf{x}_t,\mathbf{w}_t;\hat{\boldsymbol{\psi}})$. This has an adverse effect on the empirical size of the test. To circumvent this problem, we regress the residuals $\hat{u}_t$ on $\hat{\mathbf{h}}_t$ and compute the sum of squared residuals $SSR_0 = \sum_{t=1}^T \tilde{u}_t^2$. The new residuals $\tilde{u}_t$ are orthogonal to $\hat{\mathbf{h}}_t$.
(2) Regress $\hat{\boldsymbol{\nu}}_t$ on $\hat{\mathbf{h}}_t$ and compute the residuals $\mathbf{r}_t$.
(3) Regress a vector of ones on $\tilde{\varepsilon}_t\mathbf{r}_t$ and calculate the sum of squared residuals SSR_1.
(4) The value of the test statistic is given by

$$LM_{\chi^2}^r = T - SSR_1. \quad (12)$$

Under H_0, $LM_{\chi^2}^{hn}$ has an asymptotic χ^2 distribution with m degrees of freedom.

[3] See Teräsvirta (1994) for the technical conditions for the validity of the test statistic.

3. Empirical results

In this section we discuss how different specifications of the STR-Tree model actually describe the realized volatility series of DJIA stocks. Are there statistically significant structural breaks and/or regime shifts? What are the magnitudes and durations of those regimes? Are the level changes economically relevant? What does the estimation of structural breaks say about the stock market in the period? What are the in-sample fitting and out-of-sample forecasting properties of these models in relation to alternative models, such as the ARFIMA and GARCH models?

The empirical analysis focuses on the realized volatility of the sixteen Dow Jones Industrial Average index stocks that were available to us: Alcoa (AA), American International Group (AIG), Boeing (BA), Caterpillar (CAT), General Electric (GE), General Motors (GM), Hewlett Packard (HP), IBM, Intel (INTC), Johnson and Johnson (JNJ), Coca-Cola (KO), Merk (MRK), Microsoft (MSFT), Pfizer (PFE), Wal-Mart (WMT), and Exxon (XON). The raw intraday data consist of tick-by-tick quotes extracted from the NYSE Trade and Quote (TAQ) database. The period of analysis starts in January 3, 1994, and ends in December 31, 2003. Trading days with an abnormally small trading volume and volatility caused by the proximity of holidays (for example, Good Friday) are excluded, leaving a total of 2541 daily observations.

We start by removing non-standard quotes, computing mid-quote prices, filtering possible errors, and obtaining one-second returns for the 9:30 am to 16:05 pm period. Following the results of Hansen and Lunde (2006b), we adopt the *previous tick* method for determining prices at precise time marks. Based on the results of Hasbrouck (1995), who reports a median 92.7% information share at the NYSE for Dow Jones stocks, and Blume and Goldstein (1997), who conclude that NYSE quotes match or determine the best displayed quote most of the time, we use NYSE quotes (or NASDAQ, for Microsoft and Intel) if they are close enough to the time marks in relation to other updates.

In order to estimate our measure of the daily realized volatility, we use the two time scales estimator of Zhang, Mykland, and Aït-Sahalia (2005),[4] which is among the most accurate estimators available, despite requiring possibly unrealistic assumptions about the noise for consistency. The final dependent variable is the daily logarithm of the realized volatility. We also consider dummies for the days of the week, as done by Martens, van Dijk, and de Pooter (this issue),[5] and dummies for the following macroeconomic announcements: Federal Open Market Committee meetings (FOM), the Employment Situation Report from the Bureau of Labor Statistics (ESR), CPI and PPI.

Concerning the macroeconomic announcements, we have the following comments. It is not our goal to estimate possible causal effects of announcements on volatility. Of course, "good" announcements may have a different impact from "bad" announcements, although we suspect that this asymmetry affects returns more than volatility. However, for all of the series considered in this paper, the inclusion of a single dummy variable for each announcement is sufficient to smooth the effect of possible jumps in the volatility process, which is our purpose in including this type of variable.[6] In Section 3.3.2 we show that explicitly taking into account the presence of jumps does not bring any significant benefit in terms of forecasting performance. Furthermore, the macroeconomic announcements included in the model are the most relevant ones. Finally, including dummy variables for all possible announcements and all possible outcomes will heavily over-parameterize the model, which is far beyond the scope of this paper.

In Section 3.1 we present the modeling cycle adopted in the empirical experiment. We carefully discuss variable selection and model specification. In order to evaluate the benefits of the STR-Tree model over standard models, we conduct a full sample study in Section 3.2, using data from 1994 to 2003. The goal of this analysis is to point out how the STR-Tree models may be useful for describing interesting stylized facts of financial time series such as long range dependence and asymmetries. We highlight our results for the particular case of the IBM stock. The

[4] With five-minute grids in the slow time scale and one-second returns on the fast time scale.

[5] This is a standard practice in the literature.

[6] Another possibility is to filter the jumps using an estimator such as the one proposed by Barndorff-Nielsen and Shephard (2006). However, this estimator is not robust to general microstructure noise.

results for the other 15 stocks are fairly similar, and will be omitted for the sake of conciseness. Four versions of the STR-Tree model are estimated: a pure structural break model (STR-Tree/SB), where time is the single transition variable; an asymmetric effects model (STR-Tree/AE), where past cumulated returns of the stock over different horizons (reflecting the "long-run" dynamics of the market) are the candidates for controlling the regime switches; an asymmetric effects model (STR-Tree/DJIA), where past cumulated returns of the DJIA index are used as transition variables; and finally, a model combining structural breaks and asymmetric effects (STR-Tree/AE+SB), where both time and past cumulated returns are considered as split variables. We show that the asymmetric effects model successfully describes the long range dependence in the volatility of the stocks. Furthermore, using market returns (DJIA) or firm-specific returns causes no important differences in terms of in-sample performance. The in-sample results are compared with the linear ARFIMA model and the Heterogenous Autoregressive (HAR) model put forward by Corsi (2004).

In Section 3.3 we conduct an out-of-sample forecasting experiment, considering the last four years of the sample, from January 3, 2000 to December 31, 2003, covering 983 days. Each model is re-estimated daily using the full sample up to that date, and then used for out-of-sample forecasting for the horizons of 1, 5, 10 and 20 days ahead. The specifications of the STR-Tree models are revised monthly. Point forecasts for the nonlinear models are calculated through conditional simulation,[7] together with interval forecasts for all models. As reference models, we also include predictions from linear autoregressive (AR), GARCH(1,1)[8] and exponentially weighted moving average (EWMA) models. With respect to the last, we take a different approach from the literature and compute an EWMA on the realized volatility itself. The STR-Tree/DJIA is not used to compute forecasts more than one day ahead, due the non-availability of the realized variance series for the index, which is essential in the conditional simulation.

[7] See the Appendix for details.

[8] We also considered other GARCH models such as EGARCH and GJR-GARCH; however, the performance of such models is not substantially different.

3.1. Specification

Following the specific-to-general principle, we start the cycle from the root node (depth 0). Our general basic linear equation is given by:

$$\begin{aligned}\log(RV_t) &= \alpha_1 \log(RV_{t-1}) + \cdots + \alpha_k \log(RV_{t-k}) \\ &\quad + \delta_1 I[\text{Mon}]_t + \delta_2 I[\text{Tue}]_t + \delta_3 I[\text{Wed}]_t \\ &\quad + \delta_4 I[\text{Thu}]_t + \delta_5 I[\text{Fri}]_t + \delta_6 I[\text{FOMC}]_t \\ &\quad + \delta_7 I[\text{EMP}]_t + \delta_8 I[\text{CPI}]_t + \delta_9 I[\text{PPI}]_t + \varepsilon_t,\end{aligned} \tag{13}$$

where $I[\text{Mon}]_t$, $I[\text{Tue}]_t$, $I[\text{Wed}]_t$, $I[\text{Thu}]_t$, and $I[\text{Fri}]_t$ are days-of-the-week dummies and $I[\text{FOMC}]_t$, $I[\text{EMP}]_t$, $I[\text{CPI}]_t$, and $I[\text{PPI}]_t$ are dummies indicating dates for the following macroeconomic announcements: Federal Open Market Committee meetings, the Employment Situation report, CPI and PPI. Some authors discuss the relationship between macroeconomic announcements and jumps; see, for example, Barndorff-Nielsen and Shephard (2006) and Huang (2006).

The first step in the modeling cycle is to use Eq. (13) to select the number of autoregressive lags and relevant day-of-the-week and announcement effects (variables that will be in $\mathbf{w}_t$), resulting in the primary specification that will be contrasted with non-linearity. Autoregressive (AR) coefficients are tested up to the 15th order. Seeking a parsimonious specification, we base this selection on the Schwarz information criterion (SBIC), which initially selects autoregressive lags 1–3, 5, 7, 10 for all stocks, and keeps the Monday dummy for some stocks and both the Monday and Friday dummies for others. The SBIC also selects the FOMC and EMP announcements. We verified that the inclusion of a moving average (MA) term could significantly cut down the number of AR terms, but we chose the less parsimonious AR specification, since the computational burden of estimating an MA coefficient in a nonlinear framework is high, and there are sufficient degrees of freedom. The presence of an MA coefficient could be justified by the existence of both persistent and non-persistent components in volatility, such as measurement noise or jump components.[9] We consider the importance of jump components in Section 3.3.2.

[9] See Andersen, Bollerslev, and Diebold (2005) and Tauchen and Zhou (2005).

The next step is to select the set of variables in vectors $\mathbf{x}_t$ and $\mathbf{z}_t$. Over the next sections, the candidate split variables $\mathbf{z}_t$ falls in one of three cases: structural breaks (time is the unique transition variable), asymmetric effects (lagged returns and lagged cumulated returns over the past 2 to 120 days), and finally, the combination of structural breaks and asymmetric effects. A fourth possibility, explored by Martens et al. (this issue), is the inclusion of lags of the realized volatility itself as split variables. However, this particular choice of asymmetry was not significant in any of the cases analyzed. At each node, the transition variable is selected as the one that minimizes the p-value of the robust version of the LM test.

The elements of the vector $\mathbf{z}_t$ are selected as a trade-off between parsimony/interpretability and fitting properties. In the structural break case, we include the first two lags of the logarithm of the realized volatility, such that $\mathbf{z}_t = (\log(RV_{t-1}), \log(RV_{t-2}))'$. In the asymmetric effects model we set $\mathbf{z}_t = \emptyset$, such that $\tilde{\mathbf{z}}_t$ in Eq. (3) is just a constant.[10] Diagnostic statistics for all models are shown in Table 2.

3.2. Structural breaks, regime switches and long memory: A full sample evaluation

We start by following the recent literature and examining the effects of possible structural breaks on volatility levels (see, for example, Granger & Hyung, 2004; Martens et al., this issue; and Morana & Beltratti, 2004). The final estimated model for the case of IBM is given by

$$
\begin{aligned}
\log(RV_t) &= \underset{(0.164)}{0.261}\log(RV_{t-1}) + \underset{(0.078)}{0.224}\log(RV_{t-2}) \\
&+ \underset{(0.021)}{0.084}\log(RV_{t-3}) + \underset{(0.020)}{0.074}\log(RV_{t-5}) \\
&+ \underset{(0.019)}{0.044}\log(RV_{t-7}) + \underset{(0.018)}{0.047}\log(RV_{t-10}) \\
&- \underset{(0.013)}{0.064}\, I[\text{Mon}]_t - \underset{(0.014)}{0.063}\, I[\text{Fri}]_t \\
&+ \underset{(0.032)}{0.067}\, I[\text{FOMC}]_t + \underset{(0.023)}{0.094}\, I[\text{EMP}]_t \\
&+ \left\{\underset{(0.048)}{0.005} + \underset{(0.164)}{0.261}\log(RV_{t-1}) + \underset{(0.078)}{0.224}\log(RV_{t-2})\right\} \\
&\times G\left(t;\, \underset{(6.154)}{13.359},\, \underset{(0.136)}{1.744}\right) G\left(t;\, \underset{(12.716)}{7.003},\, \underset{(0.101)}{3.273}\right) \\
&+ \left\{\underset{(0.021)}{0.140} + \underset{(0.036)}{0.449}\log(RV_{t-1}) + \underset{(0.037)}{0.156}\log(RV_{t-2})\right\} \\
&\times G\left(t;\, \underset{(6.154)}{13.359},\, \underset{(0.136)}{1.744}\right)\left[1 - G\left(t;\, \underset{(12.716)}{7.003},\, \underset{(0.101)}{3.273}\right)\right] \\
&+ \left\{\underset{(0.014)}{0.118} + \underset{(0.033)}{0.409}\log(RV_{t-1}) + \underset{(0.080)}{0.033}\log(RV_{t-2})\right\} \\
&\times \left[1 - G\left(t;\, \underset{(6.154)}{13.359},\, \underset{(0.136)}{1.744}\right)\right] \\
&\times \left[1 - G\left(t;\, \underset{(12.716)}{7.003},\, \underset{(0.101)}{3.273}\right)\right] + \hat{\varepsilon}_t.
\end{aligned}
$$

The final model has 23 estimated parameters. Although it may seem over-parameterized, we stress the fact that we have a large number of observations. Two breaks are estimated: one in August 1998 (volatility and persistence[11] go up, and the unconditional mean of the daily realized volatility goes from 1.50% to 2.10%, a 40% increase), and another one in April 2003 (volatility markedly falls, and the unconditional mean goes down from 2.10% to 1.15%, a 45% decrease). The first parameter change is rather abrupt (large value of the slope parameter γ) and the second one is quite smooth (small γ). Note that the standard errors for the slope parameter estimates are quite high. Nevertheless, this is not an indication that the nonlinear effects are not significant. Due to the identification problem previously discussed in Section 2.3, the distribution of the usual t-statistic is not standard under $\mathcal{H}_0 : \gamma = 0$. The LM test is an adequate way to assess the statistical relevance of the structural changes; see Eitrheim and Teräsvirta (1996) for a discussion.

Fig. 2 puts the timing of the breaks in context, depicting the two estimated transition functions (dotted lines), the log realized volatility for the period, and the evolution of the stock price, adjusted for dividends for the 1995–2003 period. The first break coincides exactly with the Russian Crisis in 1998, whilst the second limits two distinct dynamics for the Dow Jones Industrial Average (DJIA): while the index reaches a four year bottom by October 2002, the following year is a highly positive one for the index, which climbed 25% through the period. It also

[10] More general specifications of $\mathbf{z}_t$, while statistically significantly different, brought no important out-of-sample gains, instead excessively increasing the number of estimated parameters, and occasionally causing numerical problems in the estimation algorithm.

[11] Measured as the sum of the autoregressive parameters.

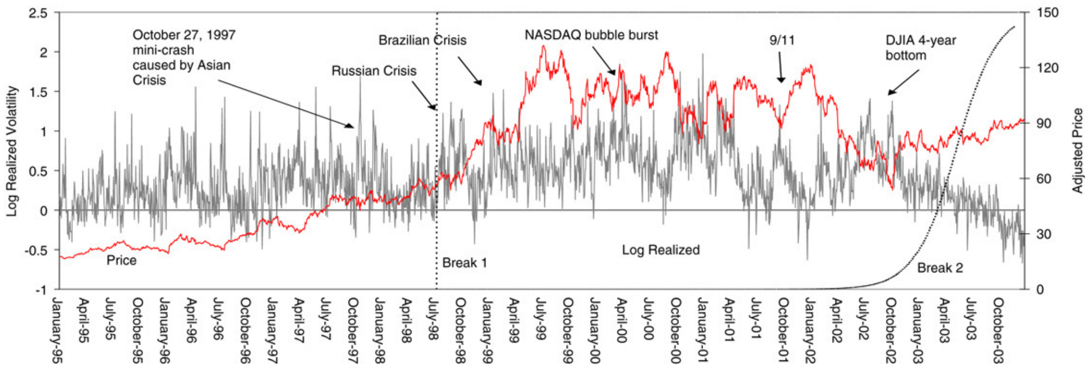

Fig. 2. IBM daily log realized volatility (1995–2003), the dividend-adjusted stock price (1995–2003), and the transition functions (dotted lines).

seems that the second break is capturing the apparent negative trend in the volatility dynamics. Fig. 2 is suggestive of other facts: there are several clusters of high volatility associated with periods of large falls in the stock price, followed by sharp declines in volatility after the price jumps up again. Some examples are the periods of the October 1997 mini-crash, the Russian crisis, the NASDAQ bubble burst, the two clusters at the end of 2000/beginning of 2001, the 9/11 period, and the bear market of 2002. The subsample between the first and second breaks (or the high volatility period) is marked by a greater incidence of these price decreases. In the next section, we turn our attention to this specific aspect.

3.2.1. Asymmetric effects

The motivation for the estimation of lagged cumulated returns as a source of multiple regimes in volatility in the STR-Tree model is illustrated in Fig. 3, which shows the realized volatility and monthly returns of IBM and the DJIA index for the period 2000 to 2003. There seems to be a recurring pattern of shifts to higher volatility levels being related to interludes of negative returns and reversals to low volatility levels in positive months. The single exception is the period before the NASDAQ bubble burst.

As mentioned before, we estimate two asymmetric effects models. In the first, past cumulated returns of the stock over different horizons are the candidates for controlling the regime switches (STR-Tree/AE), while the second has past cumulated returns of the DJIA index as transition variables (STR-Tree/DJIA). The reason for also considering Dow Jones returns is to check whether there are any substantial differences between regime switches driven by idiosyncratic factors (firm specific returns) and those driven by market factors (DJIA returns). A third model which was considered combines asymmetric effects and time changes (both past cumulated returns and time are candidate transition variables). The motivation for this is clear from Fig. 3: there is clearly a break in the volatility dynamics in 2003.

The estimated tree structure for the first model is shown in Fig. 4, and is determined by the sets $\mathbb{T} = \{1, 6, 11, 23, 24\}$ and $\mathbb{J} = \{0, 2, 5, 12\}$. The transition variables are divided by their respective standard deviations. The model is described by five highly statistically significant regimes, determined by four levels of asymmetric effects. The first node indicates a low volatility regime linked to a rising market at the four month horizon. At the other extreme, a decline of 12% or more over nearly two months induces a regime of high variance, while superior returns over this same period bring intermediate volatility levels and short run leverage effects. Negative returns over two days also induce a regime of high variance. The estimated transition functions are illustrated in Fig. 5.

The final estimated STR-Tree/AE model is given by

$$\begin{aligned}
\log(RV_t) = {} & \underset{(0.022)}{0.386}\log(RV_{t-1}) + \underset{(0.023)}{0.118}\log(RV_{t-2}) \\
& + \underset{(0.021)}{0.107}\log(RV_{t-3}) + \underset{(0.020)}{0.091}\log(RV_{t-5}) \\
& + \underset{(0.019)}{0.065}\log(RV_{t-7}) + \underset{(0.018)}{0.078}\log(RV_{t-10}) \\
& - \underset{(0.012)}{0.068}\, I[\mathrm{Mon}]_t - \underset{(0.014)}{0.064}\, I[\mathrm{Fri}]_t \\
& + \underset{(0.032)}{0.068}\, I[\mathrm{FOMC}]_t + \underset{(0.023)}{0.092}\, I[\mathrm{EMP}]_t
\end{aligned}$$

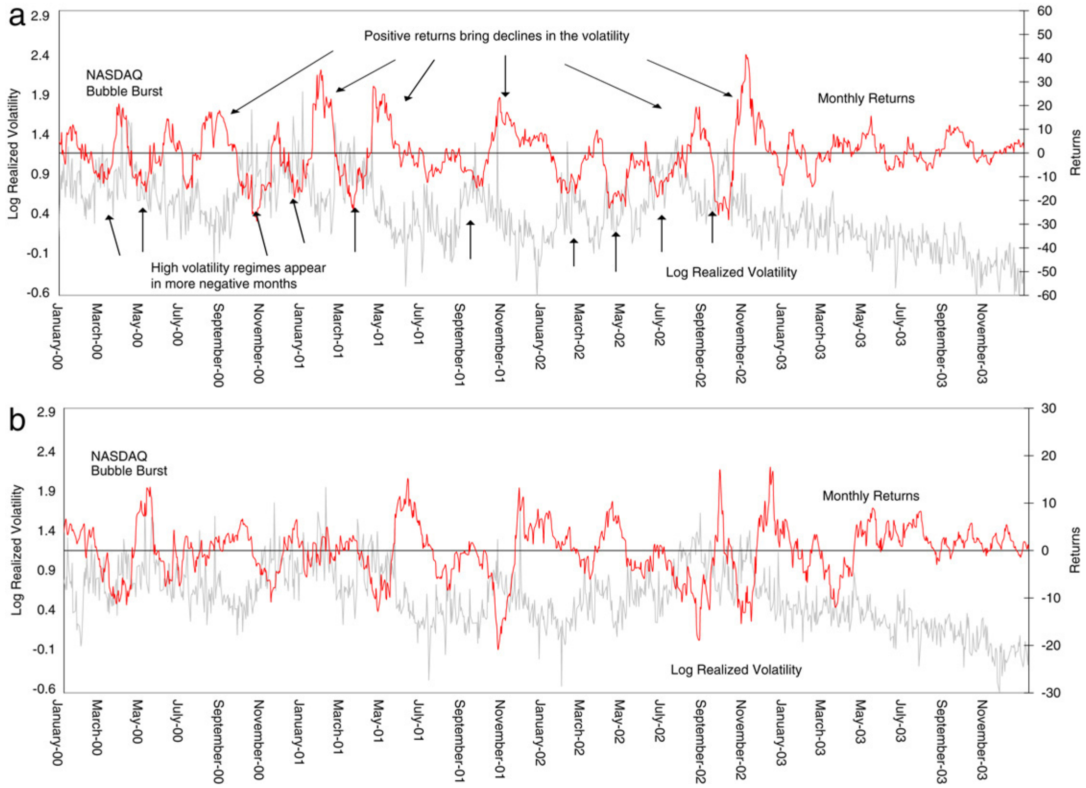

Fig. 3. Panel (a): Realized volatility and monthly IBM returns. Panel (b): Realized volatility and monthly DJIA returns.

$$
\begin{aligned}
&+ \underset{(0.013)}{0.081} \times G\left(r_{90,t-1};\ \underset{(1.082)}{2.000},\ \underset{(0.344)}{0.541}\right) \\
&+ \underset{(0.030)}{0.184} \times \left[1 - G\left(r_{90,t-1};\ \underset{(1.082)}{2.000},\ \underset{(0.344)}{0.541}\right)\right] \\
&\times \left[1 - G\left(r_{39,t-1};\ \underset{(1.018)}{2.000},\ -\underset{(0.319)}{0.955}\right)\right] \\
&- \underset{(0.046)}{0.004} \times \left[1 - G\left(r_{90,t-1};\ \underset{(1.082)}{2.000},\ \underset{(0.344)}{0.541}\right)\right] \\
&\times G\left(r_{39,t-1};\ \underset{(1.018)}{2.000},\ -\underset{(0.319)}{0.955}\right) \\
&\times G\left(r_{5,t-1};\ \underset{(1.794)}{2.000},\ \underset{(0.469)}{0.479}\right) \\
&+ \underset{(0.044)}{0.069} \times \left[1 - G\left(r_{90,t-1};\ \underset{(1.082)}{2.000},\ \underset{(0.344)}{0.541}\right)\right] \\
&\times G\left(r_{39,t-1};\ \underset{(1.018)}{2.000},\ -\underset{(0.319)}{0.955}\right) \\
&\times \left[1 - G\left(r_{5,t-1};\ \underset{(1.794)}{2.000},\ \underset{(0.469)}{0.479}\right)\right] \\
&\times G\left(r_{2,t-1};\ \underset{(1.211)}{2.423},\ -\underset{(0.284)}{1.091}\right) \\
&+ \underset{(0.127)}{0.447} \times \left[1 - G\left(r_{90,t-1};\ \underset{(1.082)}{2.000},\ \underset{(0.344)}{0.541}\right)\right] \\
&\times G\left(r_{39,t-1};\ \underset{(1.018)}{2.000},\ -\underset{(0.319)}{0.955}\right) \\
&\times \left[1 - G\left(r_{5,t-1};\ \underset{(1.794)}{2.000},\ \underset{(0.469)}{0.479}\right)\right] \\
&\times \left[1 - G\left(r_{2,t-1};\ \underset{(1.211)}{2.423},\ -\underset{(0.284)}{1.091}\right)\right] + \hat{\varepsilon}_t.
\end{aligned}
$$

Based on the estimated regimes and the transition graphs displayed in Fig. 5, we divide the observations into five different regimes. We split the observations according to the value of the transition functions (below or above 0.5). Table 1 reports the number of observations in each group and the respective mean and standard deviation of the realized volatility. Group 1 refers to the observations associated with terminal node 1 in Fig. 4. Groups 2 and 3 include

Table 1
Volatility regimes for IBM. Mean and standard deviation of the realized volatility for observations divided by a classification based on the STR-Tree/AE model with lagged cumulated returns as split variables.

Group	Mean	Standard deviation	Number of observations	Terminal nodes
1	1.57	0.54	1264	1
2	1.71	0.69	494	11
3	1.76	0.72	368	23
4	2.39	0.88	96	6
5	2.46	0.82	254	24
All	1.75	0.71	2476	–

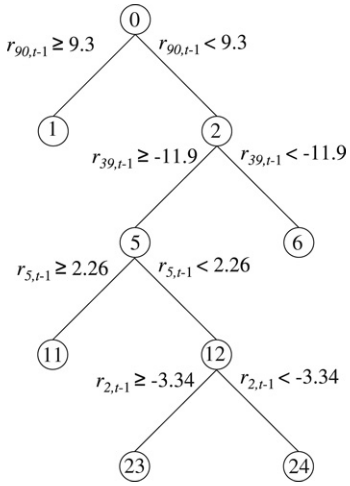

Fig. 4. Estimated tree for IBM daily log realized volatility.

observations associated with terminal nodes 11 and 23 (high returns, low volatility), respectively. Groups 4 and 5 relate to observations associated with nodes 6 and 24 (low returns, high volatility).

Finally, Fig. 6 shows the estimated functions $B_{\mathbb{J}i}(\cdot)$ for each node. Although the regime changes associated with nodes 11, 23, and 24 are quite erratic, nodes 1 and 6 induce infrequent regime changes, which, as conjectured before, may induce long range dependence.

Concerning the STR-Tree/DJIA and the STR-Tree/SB+AE, the final estimated tree architectures are described in Figs. 7 and 8. When DJIA cumulated returns are used as transition variables, the final model specification is rather different. The first split is associated with returns slightly over a month, and the subsequent splits are all associated with returns cumulated over less than a week. It seems that only the short-run dynamics of the market influence the volatility dynamics of IBM. On the other hand, when firm specific returns are considered, long-term returns appear as significant regime-switching drivers. Now we turn to the analysis of the hybrid specification, which combines both time and firm-specific cumulated returns as transition variables. First, in general the same horizons are selected as transition variables, although the location of the transition differs quite a bit. Second, two time breaks are selected, the first in November 1996 and the second in April 2003 (as in the pure break model).

3.2.2. Autoregressive fractionally integrated moving average

We now turn to the comparison of volatility models. We start with the standard ARFIMA(p, d, q), defined as

$$\phi_p(L)(1-L)^d\left[\log(RV_t)-\mu_t\right]=\theta_q(L)\varepsilon_t, \qquad (14)$$

where d denotes the fractional differencing parameter, the time-varying mean μ_t includes the day-of-the-week and announcement dates dummies, L is the lag operator, ε_t is a white noise, and $\phi_p(L)$ and $\theta_q(L)$ are polynomials of order p and q, having all roots lying outside the unit circle. For each series, we estimate several ARFIMA(p, d, q) specifications by maximum likelihood. The best combination of p, q and the dummy variables is selected by SBIC. The method leads to a choice of an ARFIMA(0, d, 0) for all series. Predictions for the ARFIMA(0, d, 0) model are computed through a truncation of the infinite autoregressive representation after the 150th lag. The final estimated model is given by:

$$(1-L)^{\underset{(0.057)}{0.516}}\left\{\log(RV_t)-\underset{(0.009)}{0.502}+\underset{(0.019)}{0.059}\,I[\mathrm{Mon}]_t + \underset{(0.020)}{0.032}\,I[\mathrm{Fri}]_t-\underset{(0.045)}{0.100}\,I[\mathrm{FOMC}]_t - \underset{(0.038)}{0.081}\,I[\mathrm{EMP}]_t\right\}=\hat{\varepsilon}_t. \qquad (15)$$

ARFIMA models have been estimated for realized volatility by Andersen et al. (2003), Beltratti and Morana (2005), Deo, Hurvich, and Lu (2006), and Martens et al. (this issue), among many others.

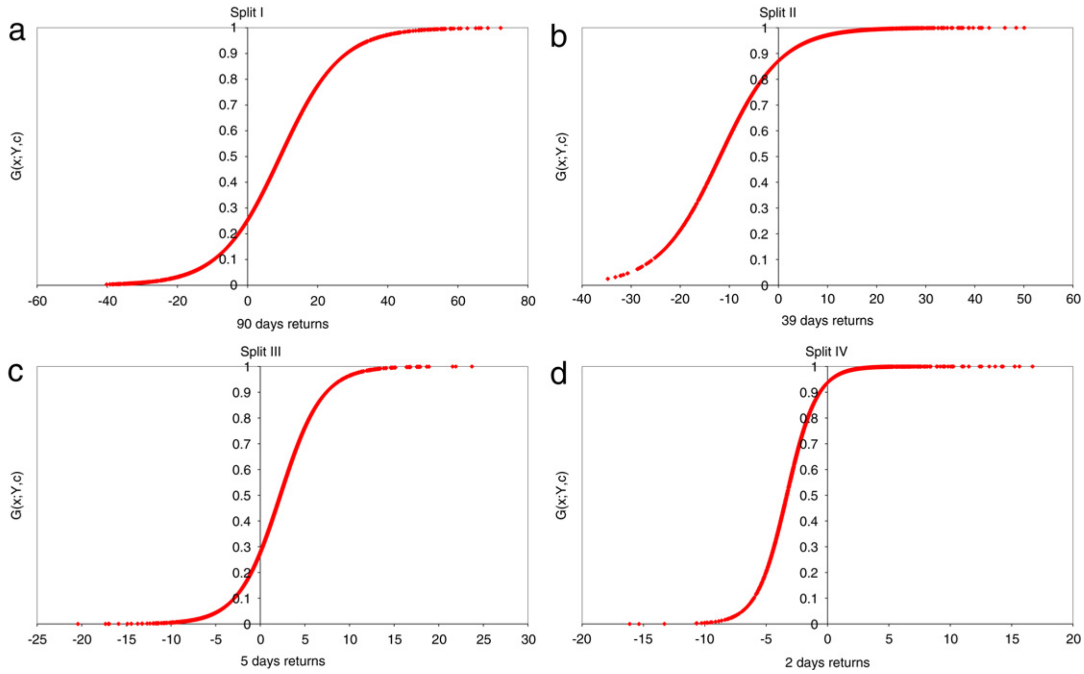

Fig. 5. Estimated transition functions.

Table 2
Estimation diagnostics. The table shows summary statistics for the residuals of six different models estimated for the log realized volatility of IBM: the STR-Tree model with lagged cumulated returns as split variables (STR-Tree/AE), the STR-Tree model with time as the split variable (STR-Tree/SB), the STR-Tree model with time and cumulated returns as transition variables (STR-Tree/SB+AE), a STR-Tree model with cumulated returns of the DJIA index as transition variables (STR-Tree/DJIA), an ARFIMA(0, d, 0) model with exogenous variables, and the HAR model. JB is the p-value of the Jarque-Bera normality test. $Q(k)$ indicates the p-value of adequate tests for serial correlation up to the kth lag. $Q^2(k)$ gives the p-value of the same tests for the squared residuals. SBIC is the Schwarz information criterion. The R^2 is corrected as was done by Andersen, Bollerslev, and Meddahi (2005).

	STR-Tree/AE	STR-Tree/SB	STR-Tree/SB+AE	STR-Tree/DJIA	ARFIMA	HAR
R^2	0.631	0.619	0.624	0.621	0.505	0.610
SD	0.223	0.226	0.225	0.225	0.255	0.229
Skewness	0.697	0.725	0.707	0.736	0.336	0.707
Kurtosis	4.703	4.535	4.780	4.815	4.166	4.503
JB	0.000	0.000	0.000	0.000	0.000	0.000
$Q(5)$	0.367	0.432	0.382	0.189	0.000	0.637
$Q(10)$	0.115	0.308	0.157	0.079	0.000	0.275
$Q(20)$	0.399	0.422	0.432	0.101	0.000	0.530
$Q^2(10)$	0.012	0.001	0.006	0.032	0.000	0.001
$Q^2(20)$	0.032	0.008	0.041	0.086	0.000	0.008
SBIC	−2.905	−2.889	−2.918	−2.919	−2.699	−2.918

3.2.3. Heterogenous autoregressive

The HAR (Heterogeneous Autoregressive) model proposed by Corsi (2004) is grounded on the Heterogeneous ARCH (HARCH) model developed by Müller et al. (1997). It is specified as a multi-component volatility model with an additive hierarchical structure, leading to an additive time series model of the realized volatility which specifies the volatility as a sum of volatility components over different horizons. The model has been used

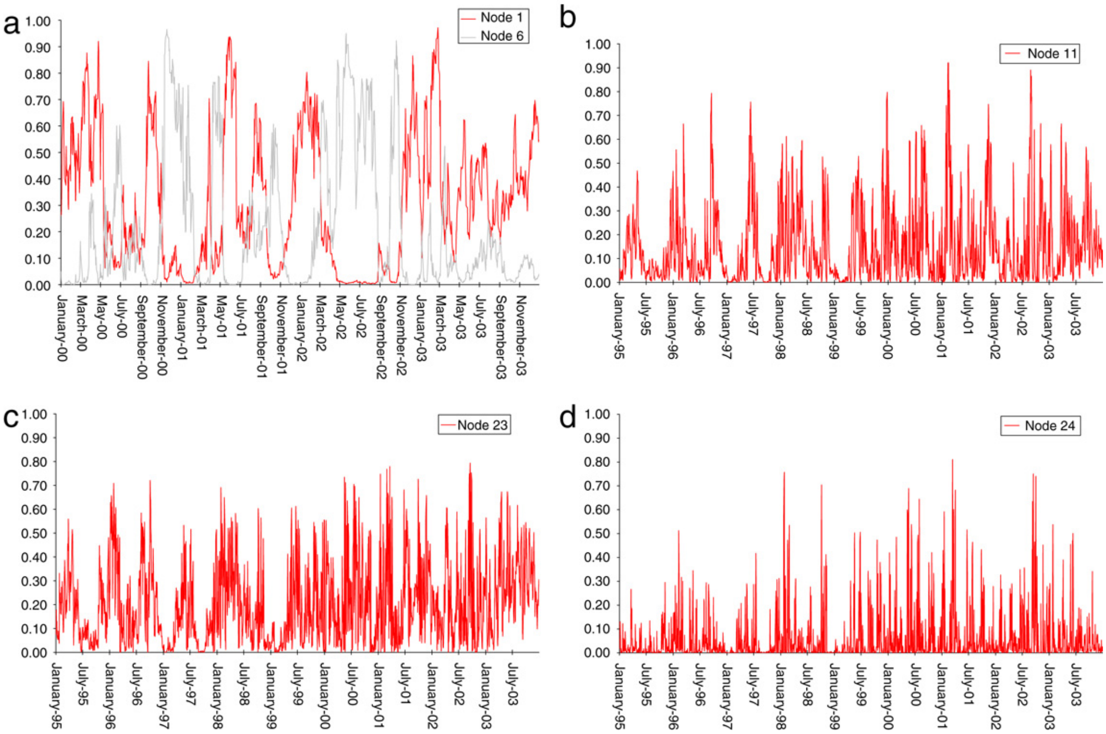

Fig. 6. Estimated regime probabilities (functions $B_{\mathbb{J}i}(\cdot)$ for each node). Panel (a) shows the estimated probabilities for nodes 1 (dark color) and 6 (light color). Panel (b) refers to node 11. Panel (c) refers to node 23. Panel (d) refers to node 24.

by Andersen, Bollerslev, and Diebold (2005), for instance, for its estimation simplicity and its capacity to reproduce the autocorrelation patterns of long memory models over shorter horizons. Define the h-horizon normalized realized volatility by

$$\log(RV_t)_{t+h} = \frac{\log(RV_{t+1}) + \log(RV_{t+2}) + \cdots + \log(RV_{t+h})}{h}. \tag{16}$$

The estimated HAR model (specified by the SBIC) is given by:

$$\begin{aligned}\log(RV_t) = &-\underset{(0.091)}{1.046} + \underset{(0.023)}{0.374}\log RV_{t-1} \\ &+ \underset{(0.026)}{0.068}\log RV_{t-2} + \underset{(0.046)}{0.247}\log(RV_t)_{t-5} \\ &+ \underset{(0.032)}{0.225}\log(RV_t)_{t-22} - \underset{(0.012)}{0.066}\, I[\text{Mon}]_t \\ &- \underset{(0.013)}{0.053}\, I[\text{Fri}]_t + \underset{(0.029)}{0.072}\, I[\text{FOMC}]_t \\ &+ \underset{(0.025)}{0.093}\, I[\text{EMP}]_t + \hat{\varepsilon}_t.\end{aligned} \tag{17}$$

We add a second order autoregressive term to the typical formulation of the model to account for remaining autocorrelation in lower lags. As before, the dummy variables are selected by minimizing the SBIC.

3.2.4. *Summary and comparison of results*

Table 2 shows summary statistics for the residuals of the four models, where JB is the p-value of the Jarque-Bera normality test, $Q(k)$ indicates the p-value of suitable tests of serial correlation up to the kth lag (the Ljung-Box portmanteau test for the ARFIMA and HAR models and a LM-type test for the nonlinear models; see Medeiros & Veiga, 2003 for a description of the latter), and $Q(k)^2$ gives the p-value of the same test for the squared residuals. The R^2 statistics are corrected according to Andersen, Bollerslev, and Meddahi (2005).

The table shows that the STR-Tree/AE model has superior in-sample fitting, as measured by R^2, while the STR-Tree/DJIA model is the best according to the SBIC. The ARFIMA model has a remarkably inferior

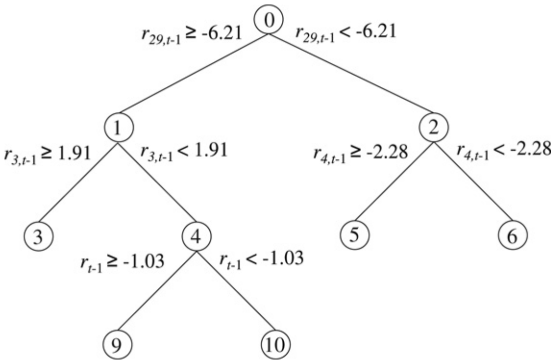

Fig. 7. Estimated tree for IBM log realized volatility with cumulated returns of the DJIA index as transition variables.

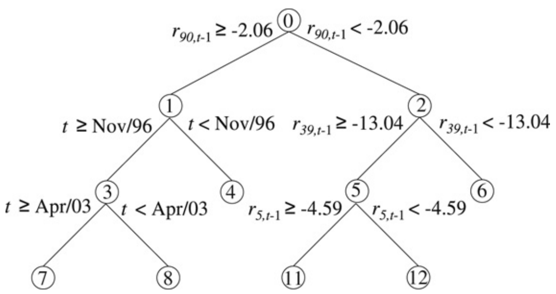

Fig. 8. Estimated tree for IBM log realized volatility with cumulated returns of IBM and time as transition variables.

fitting performance compared to the others, due to the fact that the fractional differencing parameter on the nonstationary region appears to be an inaccurate description of the series for the early years of the sample, an issue that will be discussed in more detail in the next section. All models generate highly skewed and leptokurtic residuals, indicating that the errors are not normally distributed.

The $Q(k)$ statistics in their turn indicate that only the ARFIMA model has significant remaining autocorrelation structure in the residuals up to the 20th lag at 5%. This could be due to ignored AR or MA terms in the ARFIMA, but less parsimonious models have been estimated and none of them was capable of improving on this result. Finally, there is strong evidence of dependence in squared residuals, but unlike the results of Beltratti and Morana (2005) for exchange rates, there is no indication of long memory on the conditional variance of volatility. One the other hand, all models have autocorrelated squared residuals. This is in accordance with the findings of Corsi, Mittnik, Pigorsch, and Pigorsch (2008), who model the volatility of volatility. This has an impact on the construction of the confidence intervals for the volatility forecasts, but may not influence point forecasts. The inclusion of a volatility of volatility component in the models considered here is left for future research. However, it is important to stress that in the model building strategy for the STR-Tree models, we explicitly take heteroskedasticity into account by using a robust version of the LM test.

3.2.5. Long memory analysis

To assess the long memory characteristics of the estimated STR-Tree models for IBM, we run 1000 simulations of alternative models (with the same length as the sample), and evaluate estimates of the fractional differencing parameter (d). We also include AR simulations, using the linear parameters of the STR-Tree/AE estimation to emphasize how the non-linear effects do generate hyperbolic patterns in autocorrelations beyond the possibly misleading effect of persistent autoregressive structures.

We apply two methods for the estimation of the long memory parameter: The widely used log periodogram estimator (GPH) of Geweke and Porter-Hudak (1983) and the bias reduced estimator of Andrews and Guggenberger (2003). We employ two values for the number of ordinates ℓ used in each regression: $T^{1/2}$, the usual rule of thumb value suggested by Geweke and Porter-Hudak (1983) (simulation-based), and the value selected by the plug-in method of Hurvich and Deo (1999), which points to $T^{0.65}$ for all series. T is the sample size.

For each set of simulations, we also evaluate the power of the Ohanissian, Russell, and Tsay (2004) test of true long memory process, which is based on the invariance property of the long memory parameter over temporal aggregation under the null. Andersen et al. (2001), for example, examine this property for DJIA stocks as evidence of long memory.

Table 3 reports the mean and standard deviation (in parentheses) of the fractional differencing parameter (d) estimates for the log realized volatility of IBM (entire sample), and over the simulations. The table reveals that the model with regime switching accounts for a large degree of long memory, even in large samples. In line with the literature, the same is also true for the model with structural breaks. The table

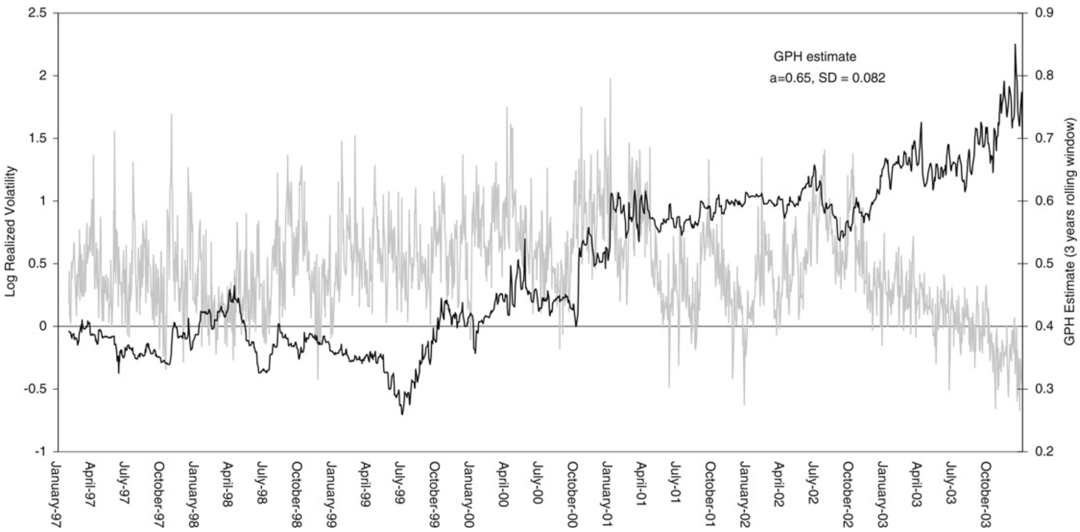

Fig. 9. GPH estimates in a rolling window.

Table 3
Log-periodogram estimates — simulations and log realized volatility. The table reports the mean and standard deviation (in parentheses) of the fractional differencing parameter (d) estimates of the IBM daily log realized volatility and over 1000 simulations of three models: the STR-Tree model with lagged cumulated returns as split variables (STR-Tree/AE), the STR-Tree model with time as the split variable (STR-Tree/SB), and the AR model. GPH and AG stand for the Geweke and Porter-Hudak (1983) and Andrews and Guggenberger (2003) estimators, respectively. The number of ordinates (ℓ) used in each regression is indicated in the first row. Two values for this parameter are employed: 0.5, the usual rule of thumb for the GPH, and 0.65, selected by the plug-in method of Hurvich and Deo (1999). The last column (OHT) gives the results for the Ohanissian et al. (2004) test of the null of a true long memory process: the first three numbers indicate the percentage of simulations where the null is rejected at the 5% level, while the last line indicates the p-value of the test for the log realized volatility of IBM.

Model	$\ell = T^{0.5}$		$\ell = T^{0.65}$		$\ell = T^{0.7}$
	GPH	AG	GPH	AG	OHT
STR-Tree/AE	0.48 (0.15)	0.30 (0.25)	0.60 (0.08)	0.44 (0.17)	33.8%
STR-Tree/SB	0.42 (0.08)	0.51 (0.12)	0.50 (0.04)	0.42 (0.09)	25.5%
AR	0.14 (0.11)	0.02 (0.16)	0.38 (0.05)	0.11 (0.11)	94.5%
Log realized vol	0.60 (0.10)	0.35 (0.17)	0.46 (0.05)	0.59 (0.10)	0.556

also shows that the Ohanissian et al. (2004) test has little power relative to these alternatives. For the log realized volatility series, the test does not reject the null hypothesis, albeit this is sensitive to the specification (ℓ and the number of aggregations) and the sample itself. For instance, if the first week is removed from the sample, the test rejects the null of long memory at the 5% level. Unfortunately, the test can almost always be tailored to favor the hypotheses of both true and spurious long memory.

An important issue with the ARFIMA approach, initially documented by Granger and Ding (1996), is the excessive variance of the fractional differencing parameter estimates over time, possibly involving extensive periods in non-stationary regions. This problem is illustrated in Fig. 9, which shows the evolution of GPH estimates ($\ell = T^{0.65}$) in a rolling window of three years over our sample. The estimates range from around 0.3 to 0.8.

An interesting feature of the STR-Tree/AE model is that it can possibly account for this fact. We illustrate this through a partial simulation of the model using the actual return series as transition variables, dividing the sample by the first estimated break in

Table 4
Log-periodogram estimates — partial simulations and log realized volatility. The table reports the mean and standard deviation (in parentheses) of the fractional differencing parameter (d) estimates of two subsamples of the daily log realized volatility of IBM and over 1000 (partial) simulations of two models: the STR-Tree model with lagged cumulated returns as split variables (STR-Tree/AE) and the STR-Tree model with time as the split variable (STR-Tree/SB). GPH and AG stand for the Geweke and Porter-Hudak (1983) and Andrews and Guggenberger (2003) estimators, respectively.

Jan/1994 to Aug/1998	GPH $\left(\ell = T^{0.5}\right)$	GPH $\left(\ell = T^{0.65}\right)$	AG $\left(\ell = T^{0.65}\right)$
STR-tree partial simulation	0.33	0.52	0.29
	(0.13)	(0.07)	(0.14)
Log realized vol	0.34	0.29	0.36
	(0.13)	(0.07)	(0.14)
Sep/1998 to Dec/2003	**GPH $\left(\ell = T^{0.5}\right)$**	**GPH $\left(\ell = T^{0.65}\right)$**	**AG $\left(\ell = T^{0.65}\right)$**
STR-tree partial simulation	0.46	0.60	0.43
	(0.11)	(0.06)	(0.12)
Log realized vol	0.65	0.66	0.74
	(0.12)	(0.07)	(0.14)

Table 5
Forecasting results. The table reports the out-of-sample forecasting results for the IBM daily realized volatility for the period between 2000 and 2003 (983 trading days, excluding days affected by holidays), where each model is re-estimated daily and used for predictions 1, 5, 10 and 20 days ahead. MAE is the mean absolute error. R^2 is the (corrected) R-squared of $RV_t = \alpha + \beta \widehat{RV}_{t|t-j,i} + \varepsilon_{t,i}$, where $\widehat{RV}_{t|t-j,i}$ is the prediction of model i for the realized volatility on day t, and RV_t is the observed realized volatility on that day. F is the p-value of the (heteroskedasticity robust) F test of the joint hypothesis that $\alpha = 0$ and $\beta = 1$. HLN is the p-value of the Harvey, Leybourne, & Newbold (1997) test of equality of the mean of loss functions, where the models are compared with the ARFIMA. SPA is the p-value of the superior predictive ability test developed by Hansen (2005). The null hypothesis is that a given model is not inferior to any of the competing models in terms of a given loss function. EWMA is the exponential weighted moving average of realized volatility itself.

	MAE	HLN	SPA	R^2	HLN	SPA	F	MAE	HLN	SPA	R^2	HLN	SPA	F
	1 day							5 days						
STR-Tree/AE	0.322	0.000	0.960	0.641	0.004	0.275	0.009	0.397	0.000	0.975	0.499	0.012	0.947	–
STR-Tree/SB	0.365	0.000	0.000	0.592	0.018	0.000	0.000	0.474	0.000	0.000	0.424	0.000	0.002	–
STR-Tree/DJIA	0.324	0.000	0.456	0.644	0.002	0.921	0.049	–	–	–	–	–	–	–
STR-Tree/SB+AE	0.340	0.485	0.004	0.610	0.304	0.011	0.938	0.409	0.185	0.285	0.495	0.071	0.841	–
HAR	0.332	0.027	0.026	0.618	0.418	0.003	0.000	0.412	0.338	0.038	0.468	0.068	0.026	–
ARFIMA	0.339	–	0.001	0.617	–	0.009	0.169	0.414	–	0.032	0.478	–	0.228	–
AR	0.334	0.092	0.001	0.616	0.497	0.004	0.000	0.410	0.215	0.020	0.467	0.066	0.021	–
EWMA	0.348	0.031	0.001	0.598	0.015	0.006	0.412	0.407	0.098	0.517	0.492	0.032	0.733	–
GARCH	0.490	0.000	0.000	0.368	0.000	0.000	0.002	0.527	0.000	0.000	0.289	0.000	0.000	–
	10 days							20 days						
STR-Tree/AE	0.447	0.003	0.969	0.388	0.048	0.878	–	0.507	0.012	0.982	0.251	0.150	0.399	–
STR-Tree/SB	0.532	0.000	0.000	0.314	0.000	0.002	–	0.604	0.000	0.000	0.172	0.000	0.002	–
STR-Tree/DJIA	–	–	–	–	–	–	–	–	–	–	–	–	–	–
STR-Tree/SB+AE	0.460	0.321	0.446	0.392	0.072	0.826	–	0.510	0.025	0.890	0.288	0.004	0.777	–
HAR	0.466	0.311	0.039	0.353	0.025	0.058	–	0.535	0.160	0.001	0.227	0.149	0.122	–
ARFIMA	0.463	–	0.287	0.370	–	0.565	–	0.524	–	0.489	0.237	–	0.269	–
AR	0.458	0.249	0.131	0.359	0.092	0.067	–	0.518	0.253	0.269	0.230	0.230	0.122	–
EWMA	0.463	0.473	0.390	0.390	0.028	0.907	–	0.536	0.090	0.233	0.252	0.107	0.370	–
GARCH	0.555	0.000	0.000	0.230	0.000	0.000	–	0.591	0.000	0.000	0.149	0.000	0.008	–

model STR-Tree/SB. Even though this simulation is ad hoc and tends to underestimate the capacity of the model to generate persistent autocorrelations, it can provide an useful indication of this ability. Table 4 shows the results, including the estimate of the log realized volatility series. As is suggested by Fig. 9, all estimates of the log realized volatility point to a significantly lower estimate for the first part of sample. In fact, this is the source of the weak in-sample performance of the ARFIMA model analyzed in Section 3.2.4—the high d estimate for the overall series produces large errors in the first subsample, as well as dependence in the residuals (which are also induced by the period of the second break). Back to the table, although the average estimates for the partial simulations are lower than the ones in the nonstationary region for the realized volatility in the second subsample, the model in fact seems to be able to reproduce this behavior.

3.3. Forecasting analysis

We base the out-of-sample analysis on the last four years of the sample, from January 3, 2000 to December 31, 2003, covering 983 days. Each model is re-estimated daily using the full sample up to that date, and then used for point forecasting for the horizons of 1, 5, 10 and 20 days ahead. The specification of the STR-Tree models is revised monthly. Point forecasts for the nonlinear models are calculated through conditional simulation, as well as interval forecasts for all models. We also include predictions generated by a GARCH(1,1) model and an exponentially weighted moving average (EWMA) model. With respect to the latter, we take a different approach from the literature, and compute an EWMA of the realized volatility.

3.3.1. Point forecasts

The evaluation of forecasts is based on the mean absolute error and the estimation of the Mincer–Zarnowitz regression[12]

$$RV_t = \alpha + \beta\widehat{RV}_{t|t-1,i} + \varepsilon_{t,i},$$

where RV_t is the observed realized volatility on day t and $\widehat{RV}_{t|t-1,i}$ is the one-step-ahead forecast of model i for the volatility on day t. If the model i is correctly specified, then $\alpha = 0$ and $\beta = 1$. We compute the (robust) p-value of the F test for this joint hypothesis and report the (corrected) R^2 of the regression as a measure of the ability of the model to track variance over time. However, the presence of heteroskedasticity hinders the computation of appropriate statistics for 5, 10, and 20 days.

We also report two tests for superior predictive ability. The first is the Harvey et al. (1997) modification of the Diebold and Mariano (1995) test of equal predictive accuracy. Each concurrent model is compared with the ARFIMA model. Let $g(e_{1t})$ and $g(e_{2t})$ denote the loss functions for the prediction errors e_{1t} and e_{2t} of models 1 and 2 on day t. For the MAE, $g(e_{it}) = |RV_t - \widehat{RV}_{t|t-j,i}|$ and for the R^2, $g(e_{it}) = [RV_t - \widehat{RV}_{t|t-j,i}]^2$. The null hypothesis is $\mathbb{E}[g(e_{1t}) - g(e_{2t})] = 0$.

The second test is the Superior Predictive Ability (SPA) test developed by Hansen (2005). The null hypothesis is that a given model is not inferior to any other competing model in terms of a given loss function.

The point forecast results are reported in Table 5. The table reports the out-of-sample forecasting results for the IBM daily realized volatility for the period between 2000 and 2003 (983 trading days). MAE is the mean absolute error. R^2 is the (corrected) R-squared value of $RV_t = \alpha + \beta\widehat{RV}_{t|t-j,i} + \varepsilon_{t,i}$, where $\widehat{RV}_{t|t-j,i}$ is the prediction of model i for the realized volatility on day t. F is the p-value of the (heteroskedasticity robust) F test of the joint hypothesis that $\alpha = 0$ and $\beta = 1$. HLN is the p-value of the Harvey, Leybourne, and Newbold (1997) test of equality of the mean of loss functions, where the models are compared with the ARFIMA. SPA is the p-value of the Superior Predictive Ability test developed by Hansen (2005). For each forecasting horizon we report two values for HLN and SPA; the first refers to the MAE and the second to the R^2 values.

For one-day-ahead forecasts, the STR-Tree/AE models are superlative in terms of both MAE and R^2, significantly outperforming the ARFIMA model (with errors 5% smaller on average), and being the only ones not rejected by the SPA tests. In the sequence, there is little distinction between the ARFIMA, AR and HAR models in terms of R^2, while the last two are slightly better in terms of MAE (the differences

[12] Hansen and Lunde (2006a) show that the R^2 of the Mincer–Zarnowitz regression ensures the correct ranking of volatility forecasts in the presence of a noisy volatility proxy. However, this is not the case with the mean absolute error, even though this problem is unlikely to be severe with the two time scales estimator of realized volatility, which appears to have a sufficiently small standard error.

are significant at the 10% and 5% levels, respectively). The model with structural breaks is markedly inferior to those alternatives. Table 8 shows the MAE for all models when one-day-ahead forecasts are considered. The values in parentheses are the p-values of the SPA test. The model with asymmetric effects has at least a small advantage for all stocks, with the exception of Pfizer (see Table 8), with the ARFIMA model being close, as the second best alternative. Although this is not shown in the paper, the STR-Tree/AE model is superior in 12 of the other series when the R^2 is considered. The ARFIMA, HAR and EWMA models alternate as the second best in terms of R^2.

The advantage of the STR-Tree/AE model in terms of the MAE is preserved when the five-day horizon is considered. The EWMA model significantly outperforms the ARFIMA model. The performances of the ARFIMA, HAR and AR models are relatively similar with respect to the MAE, with an advantage for fractional integration in R^2. The results for 10- and 20-days-ahead are similar: the STR-Tree/AE model is still the best in terms of the MAE, significantly exceeding the ARFIMA model, and being almost identical to the EWMA when R^2 is considered. However, the model with asymmetric effects and structural breaks becomes superior in R^2 for 20-day forecasts. The null hypothesis of the SPA test is no longer rejected at the 5% level for the ARFIMA, AR and EWMA specifications; HAR predictions come moderately behind.

Two surprising results that emerge from this analysis, the advantage of the AR model over the ARFIMA model in terms of MAE, even for long horizons, and the competitiveness of the simple EWMA model, highlight the difficulty of translating the long memory properties of the more advanced models into better forecasts; simple but highly persistent models already capture a good part of the predictability of the volatility series. The differences in results for the MAE and R^2 criteria, in their turn, suggest a certain degree of noise in the data. Given the large number of observations in our analysis, we see this fact as supporting the existence of structural breaks.

Back to the other stocks, Table 9 shows the MAEs for ten-day-ahead forecasts. According to the MAE, the STR-Tree/AE model is the best for twelve stocks (being the only model not to be rejected by the SPA test for at least one of the stocks), the EWMA model for two and the HAR for only one. As for the one-day-ahead forecasts, none of the ARFIMA, HAR or EWMA forecasts consistently appear as the second best. A different pattern emerges for the R^2: the EWMA model is the best for ten stocks, the STR-Tree/AE model for three, and the HAR and STR-Tree/SB+AE models each for one.

We also examine the forecasting performances of the different models by year. After 2003, the volatility consistently and sharply declined over the period, inducing autocorrelations in the residuals of all models. The results for 2000–2002 are presented in Table 6, where we concentrate on the ARFIMA and STR-Tree/AE models only. In the table, one, two or three asterisks next to MAE and/or R^2 indicate that the model has a statistically significantly lower MAE/sum of squared residuals at the 10%, 5% or 1% level, respectively, according to the Harvey et al. (1997) test.

In 2000, the STR-Tree/AE is superior for one- and five-day-ahead forecasts (significant at the 5% level), while the criteria diverge for 10 and 20 days: the ARFIMA outperforms the STR-Tree/AE in terms of the MAE, and the reverse happens with R^2. The contradiction suggests a volatility level that is unaccounted for by the STR-Tree/AE estimations, which otherwise demonstrated a superior capacity to track variations in the volatility. In 2001 and 2002, however, the STR-Tree/AE consistently and strongly outperforms the ARFIMA model over all horizons and criteria.

The statistics for 2003 are given in Table 7. For one-day-ahead forecasts, the performances of the AR, EWMA, STR-Tree/AE and HAR models are very similar, and are superior to ARFIMA, while the EWMA and HAR models have better MAEs and the ARFIMA model a higher R^2 for 20-days-ahead. The MAEs are considerably smaller than in previous years, suggesting a lower variance of the log realized volatility in the period. In fact, 20-day-ahead forecasts from the ARFIMA model have lower MAEs than the one-day-ahead forecasts for all previous years. The table also shows that the STR-Tree/SB model is strongly outperformed by ARFIMA and EWMA over the period, as in previous periods. The apparent contradiction posed by the weak performance of the break model can be seen in light of the analysis of Granger and Hyung (2004), who show that prediction with structural break models tends to be

Table 6
Forecasting results by year: 2000–2002. The table reports the out-of-sample forecasting results of the STR-Tree/AE, STR-Tree/DJIA, and ARFIMA models for each year between 2000 and 2002, where each model is re-estimated daily and used for predictions 1, 5, 10 and 20 days ahead. MAE is the mean absolute error. R^2 is the corrected R-squared value of the following regression: $RV_t = \alpha + \beta \widehat{RV}_{t,i} + \varepsilon_{t,i}$, where $\widehat{RV}_{t,i}$ is the prediction of model i for the realized volatility on day t, and RV_t is the "observed" realized volatility on that day. One, two or three asterisks next to the MAE and/or the R^2 indicate that the model has a statistically significantly lower MAE/sum of squared residuals according to the Harvey et al. (1997) test at the 10%, 5% or 1% level respectively.

	2000		2001		2002	
	MAE	R^2	MAE	R^2	MAE	R^2
	1 day					
ARFIMA	0.459	0.309	0.373	0.504	0.352	0.618
STR-Tree/AE	0.451	0.336**	0.350***	0.550***	0.328***	0.644***
STR-Tree/DJIA	0.451	0.335**	0.353***	0.556***	0.326***	0.652***
	5 days					
ARFIMA	0.536	0.129	0.465	0.390	0.454	0.357
STR-Tree/AE	0.547	0.153*	0.420***	0.405	0.428***	0.432***
STR-Tree/DJIA	–	–	–	–	–	–
	10 days					
ARFIMA	0.567***	0.082	0.537	0.233	0.525	0.190
STR-Tree/AE	0.608	0.095	0.485***	0.250	0.479***	0.288***
STR-Tree/DJIA	–	–	–	–	–	–
	20 days					
ARFIMA	0.605***	0.016	0.634	0.097	0.583	0.062
STR-Tree/AE	0.633	0.024	0.567***	0.114	0.529***	0.148***
STR-Tree/DJIA	–	–	–	–	–	–

Table 7
Forecasting results by year: 2003. The table reports the out-of-sample forecasting results for the IBM daily realized volatility for the year 2003, where each model is re-estimated daily and used for predictions 1, 5, 10 and 20 days ahead. MAE is the mean absolute error. R^2 is the (corrected) R-squared value of $RV_t = \alpha + \beta \widehat{RV}_{t|t-j,i} + \varepsilon_{t,i}$, where $\widehat{RV}_{t|t-j,i}$ is the prediction of model i for the realized volatility on day t, and RV_t is the observed realized volatility on that day. F is the p-value of the (heteroskedasticity robust) F test of the joint hypothesis that $\alpha = 0$ and $\beta = 1$. HLN is the p-value of the Harvey et al. (1997) test of equality of the mean of loss functions, where the models are compared with the ARFIMA. SPA is the p-value of the superior predictive ability test developed by Hansen (2005). The null hypothesis is that a given model is not inferior to any of the competing models in terms of a given loss function. EWMA is the exponentially weighted moving average of realized volatility itself.

	1 day							20 days						
	MAE	HLN	SPA	R^2	HLN	SPA	F	MAE	HLN	SPA	R^2	HLN	SPA	F
STR-Tree/AE	0.157	0.002	0.907	0.598	0.067	0.923	0.000	0.236	0.000	0.000	0.482	0.008	0.055	–
STR-Tree/SB	0.201	0.000	0.000	0.573	0.418	0.339	0.000	0.524	0.000	0.000	0.456	0.002	0.004	–
STR-Tree/DJIA	0.161	0.028	0.001	0.599	0.066	0.949	0.000							
STR-Tree/SB+AE	0.165	0.169	0.010	0.573	0.441	0.069	0.000	0.297	0.005	0.000	0.457	0.000	0.020	–
HAR	0.156	0.000	0.951	0.593	0.032	0.900	0.010	0.191	0.000	0.848	0.478	0.000	0.027	–
ARFIMA	0.170	–	0.000	0.569	–	0.166	0.000	0.274	–	0.000	0.546	–	0.880	–
AR	0.159	0.000	0.011	0.589	0.076	0.450	0.003	0.207	0.000	0.000	0.478	0.000	0.021	–
EWMA	0.158	0.005	0.695	0.586	0.158	0.606	0.010	0.200	0.000	0.539	0.479	0.000	0.001	–
GARCH	0.322	0.000	0.000	0.413	0.001	0.000	0.000	0.527	0.000	0.000	0.276	0.000	0.004	–

weaker even if the true process is a break process: since there is a lag in the detection of the break, moving average models perform better, a quality that is also shared by spurious ARFIMA estimations.

Table 8
One-day-ahead forecasting results for all series. The table reports the out-of-sample forecasting results (MAE) for the daily realized volatility of 15 Dow Jones stocks. The figures in parentheses are the p-values of the superior predictive ability test developed by Hansen (2005). The null hypothesis is that a given model is not inferior to any of the competing models in terms of a given loss function.

Series	STR-Tree/AE	STR-Tree/SB	STR-Tree/AE+SB	ARFIMA	HAR	EWMA
AA	0.450 (0.933)	0.476 (0.001)	0.560 (0.067)	0.456 (0.301)	0.474 (0.000)	0.465 (0.034)
AIG	0.359 (0.913)	0.371 (0.003)	0.372 (0.005)	0.364 (0.260)	0.369 (0.041)	0.371 (0.039)
BA	0.393 (0.837)	0.414 (0.002)	0.404 (0.057)	0.397 (0.469)	0.405 (0.099)	0.409 (0.063)
CAT	0.398 (0.904)	0.423 (0.000)	0.423 (0.000)	0.404 (0.152)	0.412 (0.044)	0.411 (0.063)
GE	0.340 (0.873)	0.369 (0.000)	0.363 (0.000)	0.349 (0.118)	0.355 (0.008)	0.361 (0.004)
GM	0.374 (0.920)	0.409 (0.000)	0.388 (0.003)	0.380 (0.181)	0.388 (0.003)	0.389 (0.007)
HP	0.574 (0.903)	0.604 (0.000)	0.585 (0.084)	0.579 (0.372)	0.599 (0.002)	0.583 (0.314)
INTC	0.436 (0.821)	0.490 (0.000)	0.443 (0.331)	0.448 (0.055)	0.459 (0.004)	0.466 (0.001)
JNJ	0.368 (0.806)	0.380 (0.015)	0.385 (0.008)	0.372 (0.579)	0.379 (0.139)	0.380 (0.113)
KO	0.335 (0.904)	0.360 (0.000)	0.346 (0.014)	0.341 (0.164)	0.348 (0.006)	0.339 (0.473)
MRK	0.367 (0.886)	0.389 (0.001)	0.377 (0.034)	0.370 (0.634)	0.381 (0.010)	0.378 (0.064)
MSFT	0.347 (0.827)	0.380 (0.000)	0.364 (0.000)	0.357 (0.133)	0.363 (0.013)	0.369 (0.013)
PFE	0.426 (0.036)	0.433 (0.001)	0.433 (0.002)	0.419 (0.893)	0.423 (0.531)	0.421 (0.669)
WMT	0.397 (0.882)	0.408 (0.026)	0.413 (0.008)	0.400 (0.690)	0.409 (0.036)	0.408 (0.125)
XON	0.306 (0.882)	0.322 (0.001)	0.323 (0.065)	0.312 (0.110)	0.323 (0.000)	0.321 (0.005)

3.3.2. *The effect of jumps*

So far, our analysis has not explicitly considered the presence of less persistent elements in the volatility of stocks, in contrast with the smooth and very slowly mean-reverting part associated with long memory properties. Jump components have been receiving growing amounts of attention in the realized volatility literature. Building on theoretical results for bi-power variation measures, researchers such as Andersen, Bollerslev, and Diebold (2005), Tauchen and Zhou (2005), and Barndorff-Nielsen and Shephard (2006) established related frameworks for non-parametric estimation of the jump component in asset return volatility. Empirically, Andersen, Bollerslev, and Diebold (2005) incorporate the distinction between jump and non-jump components into a forecasting model for the DM/USD exchange rate, the S&P500 market index, and the 30-year U.S. Treasury bond yield realized volatility series, and find substantial performance improvements in daily, weekly and monthly predictions.

To verify the direct impact of the jump component on our conclusions, we closely follow Andersen, Bollerslev, and Diebold (2005) and recalculate the previous forecasts using the lagged jump series as an explanatory variable for the STR-Tree/AE and HAR models.[13] The new results are displayed in Table 10. In sharp contrast to the results of Andersen, Bollerslev, and Diebold (2005), the outcome of additionally considering jumps in the realized volatility of IBM is marginal; for instance, the R^2 of daily forecasts rise from 0.641 to 0.644 and from 0.618 to 0.621 for the STR-Tree/AE and HAR models respectively.[14]

4. Conclusion

In this paper, we considered the hypothesis that cumulated price variations convey essential

[13] The bi-power variation measure necessary for computing the jump series is calculated with averaged five minute grids.

[14] However, further investigation may be needed as the jump estimators themselves become robust to microstructure noise.

Table 9
Ten-day-ahead forecasting results for all series. The table reports the out-of-sample forecasting results (MAE) for the daily realized volatility of 15 Dow Jones stocks. The figures in parentheses are the p-values of the superior predictive ability test developed by Hansen (2005). The null hypothesis is that a given model is not inferior to any of the competing models in terms of a given loss function.

Series	STR-Tree/AE	STR-Tree/SB	STR-Tree/AE+SB	ARFIMA	HAR	EWMA
AA	0.583 (0.927)	0.644 (0.003)	0.585 (0.865)	0.603 (0.064)	0.605 (0.175)	0.597 (0.368)
AIG	0.460 (0.856)	0.472 (0.249)	0.472 (0.238)	0.460 (0.874)	0.470 (0.469)	0.481 (0.122)
BA	0.512 (0.902)	0.551 (0.000)	0.530 (0.188)	0.523 (0.203)	0.515 (0.845)	0.533 (0.225)
CAT	0.499 (0.855)	0.541 (0.011)	0.520 (0.135)	0.507 (0.368)	0.517 (0.321)	0.513 (0.366)
GE	0.452 (0.884)	0.489 (0.000)	0.459 (0.616)	0.467 (0.015)	0.457 (0.776)	0.464 (0.415)
GM	0.455 (0.921)	0.526 (0.000)	0.469 (0.088)	0.481 (0.000)	0.486 (0.008)	0.483 (0.054)
HP	0.744 (0.659)	0.759 (0.327)	0.756 (0.347)	0.746 (0.560)	0.745 (0.486)	0.731 (0.727)
INTC	0.607 (0.903)	0.754 (0.000)	0.640 (0.030)	0.629 (0.014)	0.640 (0.127)	0.625 (0.408)
JNJ	0.456 (0.899)	0.477 (0.020)	0.485 (0.003)	0.460 (0.592)	0.468 (0.345)	0.487 (0.048)
KO	0.411 (0.892)	0.448 (0.001)	0.433 (0.006)	0.414 (0.574)	0.429 (0.134)	0.430 (0.121)
MRK	0.436 (0.891)	0.466 (0.000)	0.446 (0.189)	0.437 (0.753)	0.441 (0.625)	0.440 (0.541)
MSFT	0.505 (0.862)	0.551 (0.000)	0.525 (0.081)	0.510 (0.250)	0.512 (0.727)	0.520 (0.354)
PFE	0.500 (0.540)	0.536 (0.000)	0.495 (0.938)	0.508 (0.180)	0.506 (0.243)	0.510 (0.172)
WMT	0.524 (0.478)	0.536 (0.001)	0.527 (0.296)	0.518 (0.734)	0.519 (0.535)	0.511 (0.689)
XON	0.395 (0.899)	0.410 (0.099)	0.400 (0.516)	0.396 (0.797)	0.427 (0.001)	0.432 (0.000)

Table 10
Forecasting results: jumps. The table reports the out-of-sample forecasting results for the IBM volatility over the period 2000–2003 (983 trading days, excluding days affected by holidays), where each model explicitly incorporates jump components, is re-estimated daily, and is used for predictions 1, 5 and 10 days ahead. MAE is the mean absolute error. R^2 is the corrected R-squared value of the regression $RV_t = \alpha + \beta\widehat{RV}_{t,i} + \varepsilon_{t,i}$, where $\widehat{RV}_{t,i}$ is the prediction of model i for the realized volatility on day t, and RV_t is the "observed" realized volatility on that day. HLN is the p-value of the Harvey et al. (1997) test of equality of the mean of loss functions (in the table, the absolute deviation and the residuals of the regression above), where the models are compared with the ARFIMA model.

	MAE	HLN	SPA	R^2	HLN	SPA
	1 day					
STR-Tree/AE	0.324	0.000	0.340	0.644	0.001	0.785
HAR	0.334	0.079	0.001	0.621	0.259	0.004
	5 days					
STR-Tree/AE	0.398	0.000	0.793	0.500	0.005	0.968
HAR	0.410	0.198	0.004	0.472	0.194	0.007
	10 days					
STR-Tree/AE	0.450	0.008	0.504	0.386	0.068	0.742
HAR	0.463	0.480	0.014	0.355	0.033	0.041

information concerning shifts in the level of stock volatility series, and can be related to multiple regimes that induce highly persistent autocorrelations that are hard to distinguish from the patterns generated by fractionally integrated processes—even in sample sizes of several years. Using realized volatilities computed from intraday returns, we showed, not surprisingly, that volatility levels in periods of losses, like the end of 2002 (when the DJIA index reached a 4 year bottom), are significantly higher than in periods like 2003, when the index went up 25%; there is strong evidence of multiple regimes linked to return patterns in all series considered. The result was robust to the choice of firm-specific or market returns as transition variables.

We underline the importance of this analysis by presenting further evidence that fractionally integrated processes are an incomplete description of the volatility process of stocks, arguing that sharp differences in out-of-sample and in-sample performances are closely related to the empirical issue of excessive variation in estimates of the fractional differencing parameter over time.

The empirical results, in their turn, indicate that the multiple regime model proposed in the paper is a promising alternative for applications. When compared with alternative specifications with short and long memory, the more realistic model proposed in this paper is able to at least retain, and in some cases improve, the overall out-of-sample performance in forecasting. Most importantly, the model is more robust to periods of financial crises and high volatility (which are the crucial ones from the point of view of risk management), when it attains significantly better forecasts. In 15 of the 16 series considered in the paper, the STR-Tree model with past cumulated returns as transition variables is at least equivalent to, and sometimes outperforms, several concurrent models, such as the AR, ARFIMA, HAR, GARCH and EWMA models. Surprisingly, the EWMA model seems to be very competitive, especially when volatility is low, such as in 2003.

Acknowledgments

The authors are grateful to Caio Almeida, Dick van Dijk, Marcelo Fernandes, Eric Hillebrand, Michael McAleer, Les Oxley, Daniel Preve, Marco Reale, Timo Teräsvirta, Álvaro Veiga and Jun Yu for useful discussions, and to Leonardo Souza for providing ARFIMA codes and helpful comments. Financial support from CAPES and CNPq is thankfully acknowledged. Material from this paper has been presented at seminars at the Econometric Institute (Erasmus University, Rotterdam), Stockholm School of Economics, Singapore Management University, University of São Paulo, Pontifical Catholic University of Rio de Janeiro (PUC-Rio), at the XXVIII Meeting of the Brazilian Econometric Society, and at the Far Eastern Meeting of the Econometric Society 2007 (FEMES 200), Taipei, Taiwan.

Appendix. Conditional simulation

Conditional on information up to day t, forecasts for days $t+1$ through $t+k$ for the STR-Tree models are calculated through conditional simulation as follows.

(1) In the first step (one day ahead), the functional form of the model (3) is used to construct predictions conditional on past realized volatility observations and returns. Set $y_t \equiv \log(RV_t)$ and $\hat{y}_{t+k|t} \equiv \mathbb{E}\left[y_{t+k}|\mathcal{F}_t\right]$, where $\mathcal{F}_t$ is the σ-algebra of all information up to time t. Hence,

$$\hat{y}_{t+1|t} = \hat{\boldsymbol{\alpha}}'\mathbf{w}_t + \sum_{i\in\mathbb{T}} \hat{\boldsymbol{\beta}}_i \tilde{\mathbf{z}}_t B_{\mathbb{J}i}\left(\mathbf{x}_t; \hat{\boldsymbol{\theta}}_i\right).$$

(2) In order to compute multi-step forecasts, we randomly generate 10,000 NID errors ($\varepsilon_{t+1,j}$, $j = 1, \ldots, 10{,}0000$), which, when added to $\hat{y}_{t+1|t}$, make up a vector of simulated realized volatilities for day $t+1$. Even though Section 3.2 points to the inaccuracy of the normality assumption, residual bootstrapping and fat-tailed distributions do not improve the results. Under the hypothesis that standardized returns are normally distributed, we employ each of these simulated volatilities to simulate correspondent (cumulated) returns, and use Eq. (3) to compute the forecasts.

(3) Simulations are repeated for each step, generating ten thousand simulated volatility paths conditional on information up to t. The averages of these paths at each step yield the final forecasts.

References

Andersen, T., Bollerslev, T., & Diebold, F. X. (2005). *Roughing it up: Including jump components in the measurement modeling and forecasting of return volatility*. NBER working papers 11775. National Bureau of Economic Research.

Andersen, T., Bollerslev, T., & Meddahi, N. (2005). Correcting the errors: A note on volatility forecast evaluation based on high-frequency data and realized volatilities. *Econometrica*, *73*(1), 279–296.

Andersen, T., Bollerslev, T., Diebold, F. X., & Ebens, H. (2001). The distribution of realized stock return volatility. *Journal of Financial Economics*, *61*, 43–76.

Andersen, T., Bollerslev, T., Diebold, F. X., & Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, *71*, 579–625.

Andrews, D., & Guggenberger, P. (2003). A bias-reduced log-periodogram regression estimator of the long memory parameter. *Econometrica*, *71*, 675–712.

Barndorff-Nielsen, O., & Shephard, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society, Series B*, *64*, 253–280.

Barndorff-Nielsen, O., & Shephard, N. (2006). Econometrics of testing for jumps in financial economics using bipower variation. *Journal of Financial Econometrics*, *4*, 1–30.

Beltratti, A., & Morana, C. (2005). Statistical benefits of value-at-risk with long memory. *Journal of Risk*, *7*(4), 47–73.

Biais, B., Glosten, L., & Spatt, C. (2005). Market microstructure: A survey of microfoundations, empirical results, and policy implications. *Journal of Financial Markets*, *8*, 217–264.

Blume, M., & Goldstein, M. (1997). Quotes, order flow, and price discovery. *Journal of Finance*, *52*, 221–244.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *21*, 307–328.

Bollerslev, T., Litvinova, J., & Tauchen, G. (2005). *Leverage and volatility feedback effects in high-frequency data*. Working paper. Duke University.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. (1984). *Classification and regression trees*. Discussion paper. Belmont, CA: Wadsworth.

Corsi, F. (2004). *A simple long memory model of realized volatility*. Manuscript. University of Southern Switzerland.

Corsi, F., Mittnik, S., Pigorsch, C., & Pigorsch, U. (2008). The volatility of realized volatility. *Econometric Reviews*, *27*(1), 46–78.

da Rosa, J., Veiga, A., & Medeiros, M. C. (2008). Tree-structured smooth transition regression models. *Computational Statistics and Data Analysis*, *52*, 2469–2488.

Davies, R. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, *64*, 247–254.

Davies, R. (1987). Hypothesis testing when a parameter is present only under the alternative. *Biometrika*, *74*, 33–43.

de Lima, P., & Crato, N. (1993). Long range dependence in the conditional variance of stock returns. Discussion paper. In *Proceedings of the business and economic statistics section joint statistical meetings*.

Deo, R., Hurvich, C., & Lu, Y. (2006). Forecasting realized volatility using a long-memory stochastic volatility. *Journal of Econometrics*, *131*, 29–58.

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

Ding, Z., Granger, C., & Engle, R. (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, *1*, 83–106.

Eitrheim, O., & Teräsvirta, T. (1996). Testing the adequacy of smooth transition autoregressive models. *Journal of Econometrics*, *74*, 59–75.

Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, *50*, 987–1007.

Geweke, J., & Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis*, *4*, 221–238.

Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance*, *48*(5), 1779–1801.

Granger, C., & Ding, Z. (1996). Varieties of long memory models. *Journal of Econometrics*, *73*, 61–77.

Granger, C. W. J., & Hyung, N. (2004). Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns. *Journal of Empirical Finance*, *11*(3), 399–421.

Hansen, P. (2005). A test for superior predictive ability. *Journal of Business and Economic Statistics*, *23*(4), 365–380.

Hansen, P., & Lunde, A. (2006a). Consistent ranking of volatility models. *Journal of Econometrics*, *131*, 97–121.

Hansen, P. R., & Lunde, A. (2006b). Realized variance and market microstructure noise (with discussion). *Journal of Business and Economic Statistics*, *24*, 127–218.

Harvey, D., Leybourne, S., & Newbold, P. (1997). Testing the equality of mean squared errors. *International Journal of Forecasting*, *13*, 281–291.

Hasbrouck, J. (1995). One security, many markets: Determining the contributions to price discovery. *Journal of Finance*, *50*, 1175–1198.

Hillebrand, E., & Medeiros, M. (2008). *Asymmetries, breaks and long-range dependence: An estimation framework for time series of daily conditional volatility*. Discussion paper. Pontifical Catholic University of Rio de Janeiro.

Huang, X. (2006). *Macroeconomic news announcements financial market volatility and jumps*. Working paper. Duke University.

Hurvich, C., & Deo, R. (1999). Plug-in selection of the number of frequencies in regression estimates of the memory parameter of a long-memory time series. *Journal of Time Series Analysis*, *20*, 331–341.

Luukkonen, R., Saikkonen, P., & Teräsvirta, T. (1988). Testing linearity against smooth transition autoregressive models. *Biometrika*, *75*, 491–499.

Martens, M., van Dijk, D., & de Pooter, M. (2009). Forecasting S&P 500 volatility: Long memory, level shifts, leverage effects, day-of-the-week seasonality, and macroeconomic announcements. *International Journal of Forecasting*, this issue (doi:10.1016/j.ijforecast.2009.01.010).

McAleer, M., & Medeiros, M. (2008). Realized volatility: A review. *Econometric Reviews*, *27*(1), 10–45.

Medeiros, M., & Veiga, A. (2003). Diagnostic checking in a flexible nonlinear time series model. *Journal of Time Series Analysis*, *24*(4), 461–482.

Medeiros, M., & Veiga, A. (2009). Modeling multiple regimes in financial volatility with a flexible coefficient GARCH(1,1) model. *Econometric Theory*, *25*, 117–161.

Müller, U., Dacorogna, M., Dav, R., Olsen, R., & Pictet, O. (1997). Volatilities of different time resolutions – Analysing the dynamics of market components. *Journal of Empirical Finance*, *4*, 213–239.

Morana, C., & Beltratti, A. (2004). Structural change and long range dependence in volatility of exchange rates: Either, neither or both. *Journal of Empirical Finance*, *11*(4), 629–658.

Nelson, D. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, *59*(2), 347–370.

Ohanissian, A., Russell, J., & Tsay, R. (2004). *True or spurious long memory in volatility: Does it matter for pricing options?* Working paper. Graduate School of Business, University of Chicago.

Tauchen, G., & Zhou, H. (2005). *Identifying realized jumps on financial markets*. Working paper. Department of Economics, Duke University.

Taylor, S. (1986). *Modelling financial time series*. Chichester, UK: John Wiley.

Teräsvirta, T. (1994). Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association*, *89*, 208–218.

van Dijk, D., Franses, P., & Paap, R. (2002). A nonlinear long memory model, with an application to US unemployment. *Journal of Econometrics*, *110*, 135–165.

van Dijk, D., Teräsvirta, T., & Franses, P. H. (2002). Smooth transition autoregressive models – A survey of recent developments. *Econometric Reviews*, *21*, 1–47.

Wooldridge, J. (1990). A unified approach to robust, regression-based specification tests. *Econometric Theory*, *6*(1), 17–43.

Zhang, L., Mykland, P., & Aït-Sahalia, Y. (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association*, *100*, 1394–1411.

Marcel Scharth has a M.A. degree in Economics from the Pontifical Catholic University of Rio de Janeiro and is currently following the Ph.D. track in economics at the Tinbergen Institute.

Marcelo C. Medeiros is currently associate professor at the Department of Economics at the Pontifical Catholic University of Rio de Janeiro (PUC-Rio). He has an undergraduate degree in Electrical Engineering (major in Decision Systems) from PUC-Rio and an M.Sc. and a Ph.D. in control theory, optimization and statistics from the Electrical Engineering Department also at PUC-Rio. His research interests are econometrics, financial econometrics, and the link between statistical learning theory and econometrics, with special emphasis on time series analysis and forecasting. He has published papers in top international journals in statistics, machine learning and econometrics, such as the *Journal of the American Statistical Association* (JASA), *Journal of Econometrics*, *Econometric Theory*, and *IEEE Transactions on Neural Networks*.