



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

Marcel Scharth<sup>a</sup>, Marcelo C. Medeiros<sup>b,\*</sup>

<sup>a</sup> Department of Economics, The University of Amsterdam, Netherlands

<sup>b</sup> 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.

© 2009 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.

*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.<sup>1</sup> Even

---

\* Corresponding author.

E-mail addresses: [marcelsfp@gmail.com](mailto:marcelsfp@gmail.com) (M. Scharth), [mcm@econ.puc-rio.br](mailto:mcm@econ.puc-rio.br) (M.C. Medeiros).

---

<sup>1</sup> Realized variance is defined as the sum of squared intraday returns sampled at a sufficiently high frequency, consistently

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

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

where the function  $f(\cdot)$  is unknown, and, in principle, there are no assumptions about the distribution of the random term  $\varepsilon_t$ , 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  $s_j \in \mathbb{S} = \{1, 2, \dots, q\}$ . 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 ( $\sigma_t$ ) 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  $r_{5,t-1} < 0$ .  $\varepsilon_t$  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

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

The parameter  $\gamma$ , called the *slope parameter*, controls the smoothness of the logistic function. The regression

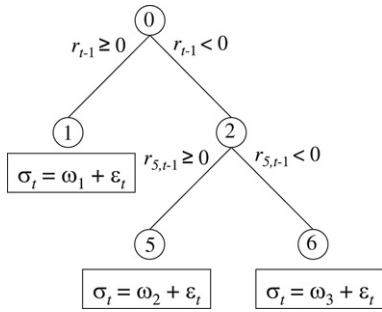


Fig. 1. Graphical representation of the volatility model described in Example 1.

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

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

where

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

and

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

where  $H_{\mathbb{J}\mathbb{T}}(\mathbf{x}_t, \mathbf{w}_t; \boldsymbol{\psi}) : \mathbb{R}^{q+1} \times \mathbb{R}^d \rightarrow \mathbb{R}$  is a nonlinear function indexed by the vector of parameters  $\boldsymbol{\psi} \in \Psi$  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  $(\gamma_k, c_k)$  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  $\sum_{i \in \mathbb{T}} B_{\mathbb{J}i}(\mathbf{x}_t; \boldsymbol{\theta}_i) = 1$ ,  $\forall \mathbf{x}_t \in \mathbb{R}^{q+1}$ .

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

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

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.<sup>2</sup>

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

<sup>2</sup> 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

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

where

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

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

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

where  $\phi_1 = \beta_{2i^*+2}$  and  $\phi_2 = \beta_{2i^*+1} - \beta_{2i^*+2}$ .

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  $\phi_2$  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

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

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

Thus, the null hypothesis becomes

$$\mathcal{H}_0 : \alpha_1 = \alpha_2 = \alpha_3 = 0. \quad (10)$$

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<sup>3</sup>:

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

where  $\hat{u}_t = y_t - H_{\mathbb{J}\mathbb{T}}(\mathbf{x}_t, \mathbf{w}_t; \hat{\psi})$ ,  $\hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^T \hat{u}_t^2$ ,  $\hat{\mathbf{h}}_t = \frac{\partial H_{\mathbb{J}\mathbb{T}}(\mathbf{x}_t, \mathbf{w}_t; \psi)'}{\partial \psi} \Big|_{\gamma_{t0}}$ , and

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

Under  $\mathcal{H}_0$ ,  $LM$  has an asymptotic  $\chi^2$  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  $H_{\mathbb{J}\mathbb{T}}(\mathbf{x}_t, \mathbf{w}_t; \hat{\psi})$ . 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  $\hat{\mathbf{v}}_t$  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  $H_0$ ,  $LM_{\chi^2}^{rn}$  has an asymptotic  $\chi^2$  distribution with  $m$  degrees of freedom.

<sup>3</sup> 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), Merck (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 Ait-Sahalia (2005),<sup>4</sup> which is

<sup>4</sup> With five-minute grids in the slow time scale and one-second returns on the fast time scale.

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),<sup>5</sup> 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.<sup>6</sup> 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

<sup>5</sup> This is a standard practice in the literature.

<sup>6</sup> 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,<sup>7</sup> together with interval forecasts for all models. As reference models, we also include predictions from linear autoregressive (AR), GARCH(1,1)<sup>8</sup> 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.

<sup>7</sup> See the Appendix for details.

<sup>8</sup> 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:

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

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.<sup>9</sup> We consider the importance of jump components in Section 3.3.2.

<sup>9</sup> 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.<sup>10</sup> 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

$$\begin{aligned} \log(RV_t) = & \underset{(0.164)}{0.261} \log(RV_{t-1}) + \underset{(0.078)}{0.224} \log(RV_{t-2}) \\ & + \underset{(0.021)}{0.084} \log(RV_{t-3}) + \underset{(0.020)}{0.074} \log(RV_{t-5}) \\ & + \underset{(0.019)}{0.044} \log(RV_{t-7}) + \underset{(0.018)}{0.047} \log(RV_{t-10}) \\ & - \underset{(0.013)}{0.064} I[\text{Mon}]_t - \underset{(0.014)}{0.063} I[\text{Fri}]_t \\ & + \underset{(0.032)}{0.067} I[\text{FOMC}]_t + \underset{(0.023)}{0.094} I[\text{EMP}]_t \\ & + \left\{ \underset{(0.048)}{0.005} + \underset{(0.164)}{0.261} \log(RV_{t-1}) + \underset{(0.078)}{0.224} \log(RV_{t-2}) \right\} \end{aligned}$$

<sup>10</sup> 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.

$$\begin{aligned} & \times G\left(t; \underset{(6.154)}{13.359}, \underset{(0.136)}{1.744}\right) G\left(t; \underset{(12.716)}{7.003}, \underset{(0.101)}{3.273}\right) \\ & + \left\{ \underset{(0.021)}{0.140} + \underset{(0.036)}{0.449} \log(RV_{t-1}) + \underset{(0.037)}{0.156} \log(RV_{t-2}) \right\} \\ & \times G\left(t; \underset{(6.154)}{13.359}, \underset{(0.136)}{1.744}\right) \left[ 1 - G\left(t; \underset{(12.716)}{7.003}, \underset{(0.101)}{3.273}\right) \right] \\ & + \left\{ \underset{(0.014)}{0.118} + \underset{(0.033)}{0.409} \log(RV_{t-1}) + \underset{(0.080)}{0.033} \log(RV_{t-2}) \right\} \\ & \times \left[ 1 - G\left(t; \underset{(6.154)}{13.359}, \underset{(0.136)}{1.744}\right) \right] \\ & \times \left[ 1 - G\left(t; \underset{(12.716)}{7.003}, \underset{(0.101)}{3.273}\right) \right] + \hat{\varepsilon}_t. \end{aligned}$$

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<sup>11</sup> 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  $\gamma$ ) and the second one is quite smooth (small  $\gamma$ ). 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

<sup>11</sup> Measured as the sum of the autoregressive parameters.



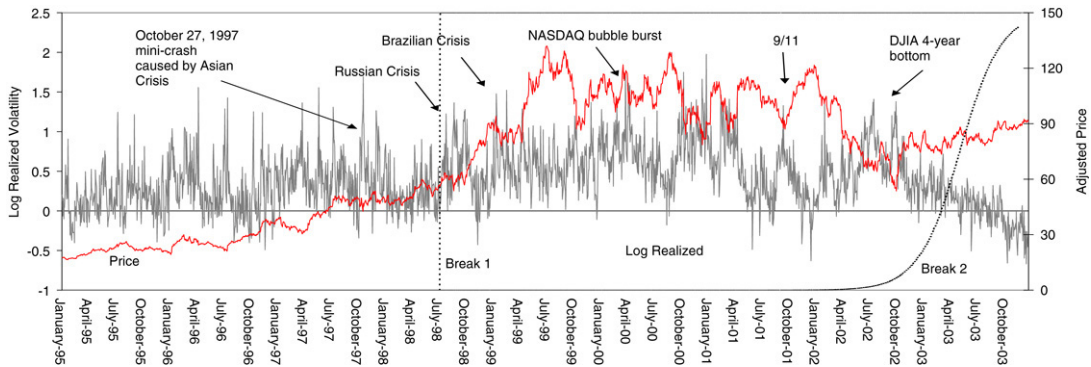


Fig. 2. IBM daily log realized volatility (1995–2003), the dividend-adjusted stock price (1995–2003), and the transition functions (dotted lines).

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

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

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

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

































