JOURNAL OF APPLIED ECONOMETRICS
J. Appl. Econ. **21**: 345–369 (2006)
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/jae.869

ESTIMATING AND PREDICTING MULTIVARIATE VOLATILITY THRESHOLDS IN GLOBAL STOCK MARKETS

FRANCESCO AUDRINO[a]* AND FABIO TROJANI[b]

[a] *University of Lugano, CH-6900 Lugano, Switzerland*
[b] *Swiss Institute of Banking and Finance, University of St. Gallen, Switzerland*

SUMMARY

We propose a general double tree structured AR-GARCH model for the analysis of global equity index returns. The model extends previous approaches by incorporating (i) several multivariate thresholds in conditional means and volatilities of index returns and (ii) a richer specification for the impact of lagged foreign (US) index returns in each threshold. We evaluate the out-of-sample forecasting power of our model for eight major equity indices in comparison to some existing volatility models in the literature. We find strong evidence for more than one multivariate threshold (more than two regimes) in conditional means and variances of global equity index returns. Such multivariate thresholds are affected by foreign (US) lagged index returns and yield a higher out-of-sample predictive power for our tree structured model setting. Copyright © 2006 John Wiley & Sons, Ltd.

1. INTRODUCTION

Volatility modelling and forecasting is one of the most important tasks in empirical finance.[1] Since the seminal work of Engle (1982) and Bollerslev (1986), several versions of a (G)ARCH (Generalized Autoregressive Conditional Heteroscedasticity) model have been proposed and widely applied in the analysis of financial markets. A limit of standard ARCH/GARCH-type models for some applications is that conditioning information (for instance on lagged return shocks) does not impact the parameters of the conditional variance equation in the model, a feature that is hardly consistent with the widely observed asymmetric behaviour of volatility in response to past positive and negative shocks.[2] Several ARCH/GARCH-type models, such as the exponential GARCH model (EGARCH; Nelson, 1991), the threshold GARCH (T-GARCH; Rabemananjara and Zakoian, 1993; Zakoian, 1994), the GJR-GARCH (Glosten *et al.*, 1993), the volatility-switching GARCH (SV-GARCH; Fornari and Mele, 1997) and the logistic smooth transition GARCH (LST-GARCH; Haregud, 1997; Gonzales-Rivera, 1998), among others, have been motivated by the attempt to incorporate asymmetric volatility dependencies. Their successful application has documented empirically the asymmetric patterns of volatility for several financial markets. It has also been noted quite recently that squared and absolute returns of financial time series typically have serial correlations that decay slowly, similarly to those of an integrated process. This pattern

* Correspondence to: Francesco Audrino, Institute of Finance, University of Lugano, Via Buffi 13, Centrocivico, CH-6900 Lugano, Switzerland. E-mail: francesco.audrino@lu.unisi.ch
[1] See Poon and Granger (2003) for a recent review.
[2] Black (1976) offers an economic explanation of the *leverage effect*, i.e. the tendence of volatility to grow and fall in response to bad and good (excess returns) news. See also the introduction in Beckaert and Wu (2000) for a description of the relation between the *risk premium effect*, i.e. the positive intertemporal relation between expected returns and conditional variances, and volatility feedbacks.

Received 3 February 2003
Revised 4 February 2005

is inconsistent with the exponential decay of conditional variance autocorrelations in a GARCH-type model and has motivated a further class of 'long memory' models for volatility. In such models, volatility shocks can impact on future volatility over very long horizons; a widely applied long memory model for volatility is, for instance, the fractionally integrated FIGARCH(1,d,1) model proposed by Baillie *et al.* (1996).[3]

This paper proposes a tree structured AR-GARCH model (Audrino and Bühlmann, 2001) for the analysis of index return series in global equity markets. The model incorporates a potentially high number of multivariate thresholds in the conditional means and volatilities of stock index returns. It is based on a binary tree structure for the definition of the thresholds. Every terminal node in the tree parameterizes a (local) AR-GARCH model for a given partition cell of a multivariate state space. The multivariate specification of the partitioning cells allows us to take into account settings where both domestic and foreign lagged index returns and volatilities can affect the conditional mean and volatility thresholds of a univariate index return series. This is a characterizing feature of our model, which is necessary to describe carefully the asymmetric patterns and the dependency of volatility in global stock markets. We propose a simple, flexible algorithm to estimate the local AR-GARCH structures in the model, as well as the number and structure of the thresholds. The optimal tree structure is identified by solving a high-dimensional model selection problem based on the Akaike Information Criterion[4] (AIC). Our model encompasses several asymmetric volatility models in the literature, as for instance the GJR-AR-GARCH model, several versions of a double TAR-GARCH model (see Li and Li, 1996; Liu *et al*., 1997) or the VS-GARCH model. Furthermore, the regime-based structure of tree structured AR-GARCH models makes them also able to produce long memory patterns in second moments, as highlighted in Audrino and Bühlmann (2001).[5]

Time series models incorporating multiple thresholds in either conditional means or volatilities are rare. Data-driven generalizations of two-regime LST-GARCH models with multiple volatility thresholds have been studied in Verhoeven and McAleer (2001) and Medeiros and Veiga (2002), who found evidence of multiple volatility regimes for six out of nine major stock indices. By contrast with our model, those papers cannot incorporate foreign lagged returns and past domestic volatilities in the definition of the thresholds. More recently, Chen *et al.* (2003) proposed a simple way of incorporating domestic and foreign lagged index returns (in their case US index returns) in a single-threshold double TAR-GARCH model, via an estimated weighted average of lagged domestic and US index returns. In our model, lagged US index returns are incorporated in some multivariate thresholds where they can interact with domestic index returns and volatilities in a quite general way. Moreover, the number of thresholds in the model is jointly estimated from the data and is not fixed from the beginning. The inclusion of lagged foreign (US) index returns in our threshold definitions is strongly motivated by all the multi-country studies collecting evidence that international stocks markets are significantly correlated, with spillovers of index returns volatility between the world's major trading areas.[6] The empirical fact that international investors often

[3] See also Hwang and Satchell (1998) and Granger (2001) for a weakness of fractionally integrated models as theoretically sound models for volatility.

[4] We use AIC because of its good overall in-sample and out-of-sample performance. However, AIC could be replaced by any sensible model selection criterion, such as for example the Schwarz Bayesian Information Criterion (BIC).

[5] See also Granger and Hyung (2000) and Diebold and Inoue (2001) for examples of short-memory models with occasional breaks or models with stochastic regimes that can exhibit long memory patterns.

[6] See, for instance, Engle *et al.* (1990), Hamao *et al.* (1990), King and Wadhwani (1990), Bae and Karolyi (1994), Kim and Rogers (1995), Koutmos and Booth (1995), Chiang (1998).

over-react to US index returns shocks, while they are less sensitive to other markets (see, for instance, Becker *et al*., 1995), supports further the hypothesis that volatility persistence in domestic equity markets can depend strongly on US equity market conditions.

We estimate our model for eight major stock indices and find strong evidence for more than one multivariate threshold (two multivariate regimes) in conditional means and variances of index returns. In particular, conditioning information from the US market affects the estimated thresholds and has strong out-of-sample predictive power. By contrast, information on past domestic volatilities does not generally affect the mean and volatility thresholds. With the exception of the Italian market only, we always find at least two regimes in the data. Such regimes are determined by an asymmetry of conditional means and volatilities as functions of 'bad' and 'good' lagged domestic index returns. In most cases we also identify one further threshold. Such a threshold is associated with the different impact of domestic index returns in the joint presence of either 'bad' or 'good' past US index returns. An interesting finding of our analysis is that the number and structure of the estimated thresholds differ across geographic areas and depend strongly on whether US index returns have been incorporated into the model.

The plan of the paper is as follows. Section 2 presents our model and the corresponding estimation procedure. The empirical results for eight stock indices of developed equity markets are presented in Section 3. Section 4 summarizes and concludes.

2. THE MODEL

This section describes our double tree-structured AR-GARCH model. In a second step, we present the algorithm and the model selection procedure that is applied to estimate it.

2.1. Starting Point

Let the daily log-return (in percentages) of a domestic stock index be denoted by $X_t = 100 * \log(P_t/P_{t-1})$, where P_t is the value of the index at day t. Similarly, we denote by X_t^{us} the US stock index return at time t, i.e. the return on the 'foreign market'. Let $\mathbf{X}_t := (X_t, X_t^{\mathrm{us}})'$ be the joint vector of domestic and foreign index returns. For exposition purposes it is useful to start from a general (nonparametric) model for X_t of the form

$$X_t = \mu_t + \varepsilon_t \tag{1}$$

where

$$\varepsilon_t = \sigma_t Z_t, \quad \mu_t = g(\mathbf{X}_{t-1}, \sigma_{t-1}^2), \quad \sigma_t^2 = f(\varepsilon_{t-1}, \mathbf{X}_{t-1}, \sigma_{t-1}^2) \tag{2}$$

for some functions $g : G = \mathbb{R}^2 \times \mathbb{R}^+ \to \mathbb{R}$ and $f : \mathbb{R} \times G \to \mathbb{R}^+$. $(Z_t)_{t\in\mathbb{Z}}$ is a sequence of iid zero mean innovations with unit variance and such that Z_t is independent of $\mathbf{X}_{t-k}$, $k = 1, \ldots, t-1$. Conditional means and volatilities are functions of lagged domestic and foreign returns, lagged domestic shocks and lagged domestic volatilities. The dependence of μ_t on σ_{t-1} and $\mathbf{X}_{t-1}$ can take into account a (possibly nonlinear) conditional mean effect of volatility and an asymmetric dependence on foreign and domestic lagged index returns. Similarly, the dependence of σ_t^2 on ε_{t-1}, σ_{t-1} and $\mathbf{X}_{t-1}$ allows for a broad variety of asymmetric volatility patterns, dependent on lagged domestic and foreign index return information $\mathbf{X}_{t-1}$.

Several models in the literature are special cases of this general setting. For instance, the GJR-GARCH models and standard T-GARCH models are encompassed by (2). Similarly, an AR-VS-GARCH(1,1) arises within (2) by setting

$$
\begin{aligned}
g(x, x^{us}, \sigma^2) &= \phi x + \psi x^{\mathrm{us}} \\
f(\varepsilon, x, x^{\mathrm{us}}, \sigma^2) &= (\alpha_{0,1} + \alpha_{1,1}\varepsilon^2 + \beta_1\sigma^2)\, I_{[\varepsilon\leq 0]} + (\alpha_{0,2} + \alpha_{1,2}\varepsilon^2 + \beta_2\sigma^2)\, I_{[\varepsilon>0]}
\end{aligned}
\tag{3}
$$

In this model, one single threshold is present in the variance equation. Moreover, it is a univariate one since it is a function only of past domestic shocks ε_{t-1}. Similarly, a double TAR-GARCH(1,1) model, as in Chen *et al*. (2003), can be written as

$$
\begin{aligned}
g(x, x^{\mathrm{us}}, \sigma^2) &= (\phi_1 wx + \phi_1(1-w)x^{\mathrm{us}})\, I_{[wx+(1-w)x^{\mathrm{us}}\leq d]} + (\phi_2 wx + \phi_2(1-w)x^{\mathrm{us}}) \\
&\quad \times I_{[wx+(1-w)x^{\mathrm{us}}>d]} \\
f(\varepsilon, x, x^{\mathrm{us}}, \sigma^2) &= (\alpha_{0,1} + \alpha_{1,1}\varepsilon^2 + \beta_1\sigma^2)\, I_{[wx+(1-w)x^{\mathrm{us}}\leq d]} + (\alpha_{0,2} + \alpha_{1,2}\varepsilon^2 + \beta_2\sigma^2) \\
&\quad \times I_{[wx+(1-w)x^{\mathrm{us}}>d]}
\end{aligned}
\tag{4}
$$

This model defines a single threshold in conditional means and volatilities. The threshold depends on domestic and foreign index returns. Therefore, lagged US market information can affect the potentially asymmetric patterns of conditional means and variances in the model. However, lagged US and domestic returns impact the threshold only through a weighted sum, which strongly constrains the model dynamics.

We propose a parametric model for (2) which admits a higher flexibility in the functional form of g and f. The model is parsimonious enough to be statistically and computationally manageable when fitting index returns based on amounts of data that are typically available in applications. We accomplish this by means of two modelling steps. First, we partition the domains of g and f into a finite number of cells. Second, for any given partition cell we specify a cell-dependent linear AR-GARCH structure for conditional means and volatilities.

2.2. Tree-Structured AR-GARCH Models

Tree-structured AR-GARCH models parameterize the conditional mean $\mu_t = \mu_t(\theta)$ and conditional volatility $\sigma_t = \sigma_t(\theta)$ in model (2) by means of some parametric threshold functions and a parameter vector θ:

$$X_t = \mu_t(\theta) + \sigma_t(\theta)Z_t \tag{5}$$

where

$$\mu_t(\theta) = g_\theta(\mathbf{X}_{t-1}, \sigma_{t-1}^2(\theta)), \quad \sigma_t^2(\theta) = f_\theta(\varepsilon_{t-1}, \mathbf{X}_{t-1}, \sigma_{t-1}^2(\theta)) \tag{6}$$

for some parametric functional forms g_θ, f_θ and $\sigma_t(\theta)$. In our tree-structured AR(1)-GARCH(1,1) model we specify $\sigma_t(\theta)$ as a threshold GARCH(1,1) function f_θ and $\mu_t(\theta)$ as a threshold linear autoregressive AR(1) function g_θ. A key property of the model is that it incorporates in the threshold definitions for f_θ and g_θ the joint impact of $\mathbf{X}_{t-1}$ and $\sigma_{t-1}^2(\theta)$. More precisely, g_θ

and f_θ are defined as threshold functions, starting from a given partition[7] $\mathcal{P}$ of the state space $G := \mathbb{R}^2 \times \mathbb{R}^+$ of $(\mathbf{X}_{t-1}, \sigma^2_{t-1}(\theta))'$:

$$\mathcal{P} = \{\mathcal{R}_1, \ldots, \mathcal{R}_k\}, \quad G = \cup_{j=1}^k \mathcal{R}_j, \quad \mathcal{R}_i \cap \mathcal{R}_j = \emptyset (i \neq j)$$

Given a partition cell $\mathcal{R}_j$, the local conditional dynamics of X_t on $\mathcal{R}_j$ are defined by an AR(1)-GARCH(1,1) model.[8] As a consequence, the threshold functions g_θ and f_θ depend on the set of parameters of any local AR(1)-GARCH(1,1) model on a partition cell $\mathcal{R}_j$, $j = 1, \ldots, k$, and the structure of the partition $\mathcal{P}$. We have:

$$\begin{aligned} g_\theta(x, x^{\mathrm{us}}, \sigma^2) &= g_\theta^{\mathcal{P}}(x, x^{\mathrm{us}}, \sigma^2) = \sum_{j=1}^{k} (\phi_j x + \psi_j x^{\mathrm{us}})\, I_{[(x, x^{\mathrm{us}}, \sigma^2) \in \mathcal{R}_j]} \\ f_\theta(\varepsilon, x, x^{\mathrm{us}}, \sigma^2) &= f_\theta^{\mathcal{P}}(\varepsilon, x, x^{\mathrm{us}}, \sigma^2) = \sum_{j=1}^{k} (\alpha_{0,j} + \alpha_{1,j}\varepsilon^2 + \beta_j \sigma^2)\, I_{[(x, x^{\mathrm{us}}, \sigma^2) \in \mathcal{R}_j]} \end{aligned} \tag{7}$$

where $\theta = (\phi_j, \psi_j, \alpha_{0,j}, \alpha_{1,j}, \beta_j; j = 1, \ldots, k)$ is a parameter vector that parameterizes the local AR-GARCH dynamics on the different partition cells $\mathcal{R}_j$, $j = 1, \ldots, k$. $k = 1$ implies a standard AR(1)-GARCH(1,1)-type model dynamics. For $k \geq 2$ we obtain a rich class of threshold models that includes for instance the AR-VS-GARCH(1,1) model (3) or the DTAR-GARCH model (4) as very special cases.

To completely specify functions f_θ and g_θ we finally define the class of partitions $\mathcal{P}$ which are admissible in a tree-structured model. Essentially, the only restriction is that $\mathcal{P}$ has to be composed of *rectangular* partition cells $\mathcal{R}_j$, $j = 1, \ldots, k$. To describe and construct such partitions for applications we make use of a binary tree where every terminal node represents a rectangular partition cell $\mathcal{R}_j$ whose edges are determined by thresholds. Figure 1 illustrates an example of a binary tree partition of the state space

$$G = \{(x, x^{\mathrm{us}}, \sigma^2); (x, x^{\mathrm{us}}) \in \mathbb{R}^2, \sigma^2 \in \mathbb{R}^+\}$$

Such a partition involves three partition cells $\mathcal{R}_1$, $\mathcal{R}_2$ and $\mathcal{R}_3$. Each partition cell $\mathcal{R}_1$, $\mathcal{R}_2$, $\mathcal{R}_3$ corresponds to a rectangular terminal node in the tree. The first cell $\mathcal{R}_1 = \{(x, x^{\mathrm{us}}, \sigma^2); x \leq d_1\}$ represents a first regime of X_t in response to 'low' lagged domestic returns. The second cell $\mathcal{R}_2 = \{(x, x^{\mathrm{us}}, \sigma^2); x > d_1 \text{ and } x^{\mathrm{us}} \leq d_2\}$ corresponds to a second regime in response to lagged 'high' domestic returns and 'low' US returns. Finally, $\mathcal{R}_3 = \{(x, x^{\mathrm{us}}, \sigma^2); x > d_1 \text{ and } x^{\mathrm{us}} > d_2\}$ represents a third regime in response to 'high' lagged domestic and US returns. It is important to remark that the threshold values d_1, d_2 in the above partition are unrestricted and are jointly estimated in our AR-GARCH tree-structured model. Thus, in applications negative lagged domestic or US returns are not constrained to imply always a relevant 'bad' market signal. Similarly, positive lagged returns do not necessarily always have to imply a 'good' market signal. 'Bad' ('good')

[7] For simplicity we present the model without including also ε_{t-1} in the partitioned state space G, even if this can be accomplished in an obvious way.

[8] The choice of local AR(1)-GARCH(1,1) models is not restrictive. The conditional dynamics in the different regimes can be more complex (for example, including also long memory in second moments). However, we find that our choice allows for a good trade-off between flexibility of the model and parsimony in the number of parameters.

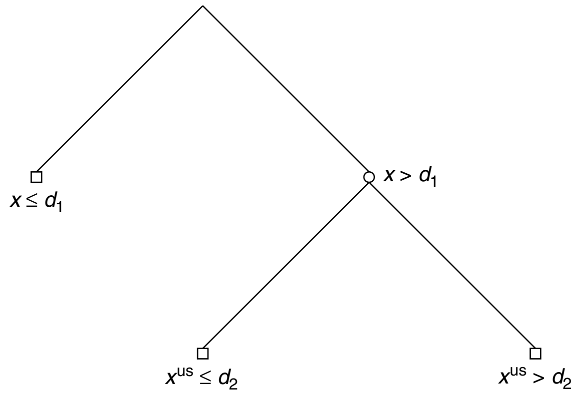

Figure 1. Example of a binary tree partition $\mathcal{P} = \{\mathcal{R}_1, \mathcal{R}_2, \mathcal{R}_3\}$ of the state space $G = \{(x, x^{us}, \sigma^2); (x, x^{us}) \in \mathbb{R}^2, \sigma^2 \in \mathbb{R}^+\}$

lagged returns with respect to the partition $\{\mathcal{R}_1, \mathcal{R}_2, \mathcal{R}_3\}$ are rather returns that are sufficiently below (above) the threshold values d_1 and d_2.

As an illustration, in our empirical application of Section 3 a partition of the form given by Figure 1 has been estimated for the SMI index returns (see Table II for more details). In this case, the estimated threshold values were $d_1 = -1.202$, $d_2 = 1.327$. Such a partition implies first an asymmetric persistence of volatility dependent on the level of lagged domestic index returns, i.e. depending on whether past SMI returns were particularly low ($x < d_1 = -1.202$). Moreover, it also highlights a particular structure of conditional SMI volatilities for states where lagged SMI returns did not realize a particularly large loss, i.e. when $x \geq d_1 = -1.202$. In these states, conditioning information about lagged US index returns becomes important to predict current SMI returns. In particular, conditioned on large lagged US index returns ($x^{us} > d_2 = 1.327$) the prevailing SMI volatility dynamics show a very different structure than in the opposite case of a 'not too large' US index return ($x^{us} \leq d_2 = 1.327$).

In the above sense, conditioning information about lagged foreign (US) index returns implies an asymmetric volatility dependence and persistence of domestic (SMI) index returns. Such an asymmetric volatility dependence and persistence could be caused, for instance, by international investors who often over-react to US index returns shocks and at the same time are less sensitive to other markets; see again Becker *et al.* (1995).

2.3. The Estimation Procedure

The negative pseudo log-likelihood[9] for model (6) is

$$-\ell(\theta; \mathbf{X}_2^n) = -\sum_{t=2}^{n} \log[\sigma_t^{-1}(\theta) p_Z((X_t - \mu_t(\theta))/\sigma_t(\theta))] \quad (8)$$

where $p_Z(\cdot)$ is a density function for the distribution of the standardized innovation Z_t and where $\mathbf{X}_2^n = \{\mathbf{X}_2, \ldots, \mathbf{X}_n\}$. Therefore, for any fixed partition $\mathcal{P}$ model (6) can be estimated by means of

[9] The log-likelihood is always considered conditionally on $\mathbf{X}_1$ and on some reasonable starting value $\sigma_1^2(\theta)$, e.g. $\sigma_1^2(\theta) = \mathrm{Var}(X_1)$.

pseudo maximum likelihood. The choice between different partition structures, i.e. the selection of the optimal threshold functions, involves a model choice procedure for non-nested hypotheses. Thus, a flexible procedure for the estimation of tree-structured AR-GARCH-type models can be based on the following two steps:

(a) For any given partition $\mathcal{P}$ the estimation of θ is performed by a pseudo maximum likelihood estimator based on a Gaussian pseudo log likelihood and on parametric forms (7) for g_θ and f_θ.
(b) Model selection of the optimal threshold function, i.e. the optimal partition $\mathcal{P}$, is performed via a tree-structured partial search.[10] Within any data-determined tree structure the optimal model is selected according to the AIC criterion.

More concretely, estimation of tree-structured AR-GARCH models by means of (a), (b) is achieved as follows. First, we estimate a largest tree-structured AR-GARCH model, given a maximal number of candidate thresholds. Second, we apply a model selection procedure for non-nested models that selects an optimal subtree of the largest tree estimated in the first step.

A parsimonious specification of the maximal number of thresholds in the first step ensures a statistically and computationally tractable model dimension. Moreover, it avoids (over)fitting a too flexible model dynamics, which would result in a poor out-of-sample forecasting power. In our applications, we fix the maximal number of candidate thresholds in the first step at 4. This implies a number of estimated parameters across the selected models including US index returns which is typically about 15 and never more than 20 in our application of Section 3.

Estimation of the 'Maximal' Tree

We fix a maximal allowed number $M+1$ of partition cells in the tree. This implies a maximal number M $(M+1)$ of possible multivariate thresholds (regimes) in conditional means and variances. For stock indices, choosing M around 4 is often appropriate (see, for example, Medeiros and Veiga, 2002). For any coordinate axis of the multivariate state space that has to be split we search for multivariate thresholds over grid points that are empirical α-quantiles of the data along the relevant coordinate axis. We fix the empirical quantiles as $\alpha = i/\text{mesh}$, $i = 1, \ldots, \text{mesh} - 1$, where mesh determines the fineness of the grid on which we search for multivariate thresholds.[11] Typically, we choose mesh = 8. The partition of the state space $G = \mathbb{R}^2 \times \mathbb{R}^+$ into a maximal number of $M+1$ cells is performed as follows. A first threshold $d_1 \in \mathbb{R}$ or $\mathbb{R}^+$ in one coordinate indexed by a component index $\iota_1 \in \{1, 2, 3\}$ partitions G as

$$G = \mathcal{R}_{left} \cup \mathcal{R}_{right}$$

where $\mathcal{R}_{left} = \{(x, x^{us}, \sigma^2) \in \mathbb{R}^2 \times \mathbb{R}^+; (x, x^{us}, \sigma^2)_{\iota_1} \leq d_1\}$ and $(x, x^{us}, \sigma^2)_{\iota_1}$ is the ι_1 component of the tuple (x, x^{us}, σ^2). $\mathcal{R}_{right}$ is defined analogously using the relation '$>$' instead of $\leq$. In a second step, one of the partition cells $\mathcal{R}_{left}$, $\mathcal{R}_{right}$ is again partitioned with a second threshold d_2 and a second component index ι_2 in the same way as above.

[10] This avoids a computationally infeasible exhaustive search.
[11] For the sample sizes available in our empirical application such a choice of the grid fineness works well, as demonstrated also by Audrino and Bühlmann (2001). A finer choice of the grid would also rapidly increase the computation time needed for our real data exercise.

We iterate this procedure. Specifically, for the mth iteration step, we specify a new pair (d_m, ι_m) (that determines a new threshold d_m for the coordinate indexed by ι_m) and an existing partition cell which is going to be further split into two subcells. For a new pair $(d, \iota) \in \mathbb{R} \times \{1, 2, 3\}$ refinement of an existing partition $\mathcal{P}^{(old)}$ is obtained by picking $\mathcal{R}_{j^*} \in \mathcal{P}^{(old)}$ and splitting it as

$$\mathcal{R}_{j^*} = \mathcal{R}_{j^*,left} \cup \mathcal{R}_{j^*,right} \tag{9}$$

This gives a new (finer) partition of G as

$$\mathcal{P}^{(new)} = \{\mathcal{R}_j, \mathcal{R}_{j^*,left}, \mathcal{R}_{j^*,right}, j \neq j^*\} \tag{10}$$

where (d, ι) describes a threshold and a component index such that $\mathcal{R}_{j^*,left} = \{(x, x^{\mathrm{us}}, \sigma^2) \in \mathcal{R}_{j^*}; (x, x^{\mathrm{us}}, \sigma^2)_\iota \leq d\}$. $\mathcal{R}_{j^*,right}$ is defined analogously, with the relation '$>$' instead of '$\leq$'. The whole procedure finally determines a partition $\mathcal{P} = \{\mathcal{R}_1, \ldots, \mathcal{R}_k\}$, which can be represented and summarized by a binary tree where every terminal node represents a partition cell in $\mathcal{P}$, see again Figure 1. To select the specific threshold and component index (d, ι) in each iteration step of the above procedure we proceed by optimizing a conditional negative (pseudo) log-likelihood (8). Details on the implied algorithm are given in Appendix A.

Proofs of consistency of the model selection procedure for the case that the true model is in the class of tree-structured AR-GARCH models are very difficult to obtain. In particular, using the model selection algorithm proposed in Appendix A, which keeps the parameters outside the refined cells fixed in step 2-II, the search is not guaranteed to end up giving the correct structure. This can in principle be corrected by a full pseudo maximum likelihood estimation in step 2-II. However, as for the classical search algorithms used for classification and regression trees (CART; Breiman *et al.*, 1984), this would imply considerable extra costs and is not computationally possible. Analogously to CART, it is possible to prove theorems that study the behaviour of the prevailing parameter estimators when growing the tree. However, such results also do not imply model selection consistency. Furthermore, it is quite hard to believe that the 'correct' generating process in our and similar real data examples is indeed a tree-structured AR-GARCH-type model. For this reason, it is more important to prove consistency of the volatility estimates in a tree-structured GARCH model under a possible model mis-specification, rather than showing consistency of the model selection strategy under the assumption of a correctly specified tree-structured model. Such consistency results can be found in Audrino and Bühlmann (2001).

Selection of the 'Optimal' Subtree

The maximal binary tree (or equivalently the maximal partition $\mathcal{P}_{opt}^{(M)}$) constructed with the algorithm presented in Appendix A can be too large (or too fine, respectively). We correct by pruning. Specifically, we search for a best subtree of $\mathcal{P}_{opt}^{(M)}$ using the AIC model selection criterion (11) below.

Let τ be the set of all binary subtrees of $\mathcal{P}_{opt}^{(M)}$ and denote for brevity an arbitrary element of τ by $\mathcal{P}_i$. Thus, every subtree $\mathcal{P}_i$ corresponds to a partition of G which can be represented as a binary tree. Note that τ is in general larger than the set of partitions $\{\mathcal{P}_{opt}^{(0)}, \mathcal{P}_{opt}^{(1)}, \ldots, \mathcal{P}_{opt}^{(M)}\}$. For every $\mathcal{P}_i$ we compute the implied pseudo maximum likelihood estimate $\hat{\theta}^{\mathcal{P}_i}$, according to (8) and based on function $g^{\mathcal{P}_i}(\cdot, \cdot, \cdot)$, $f^{\mathcal{P}_i}(\cdot, \cdot, \cdot, \cdot)$ of the form (7). We then consider the penalized negative log-likelihood (or AIC) statistic

$$\mathrm{AIC}(\mathcal{P}_i) = -2 * \ell(\hat{\theta}^{\mathcal{P}_i}; \mathbf{X}_2^n) + 2 * \dim(\hat{\theta}^{\mathcal{P}_i}) \tag{11}$$

as a measure of predictive performance for the given partition $\mathcal{P}_i$. We finally select the binary tree (or equivalently the partition) $\hat{\mathcal{P}}$ that minimizes (11) across all subtrees of the maximal tree $\mathcal{P}_{opt}^{(M)}$. The resulting estimated tree-structured AR-GARCH model is given by a parametric functional form (6) with functions $g_{\hat{\theta}^{\hat{\mathcal{P}}}}^{\hat{\mathcal{P}}}(\cdot, \cdot, \cdot)$ and $f_{\hat{\theta}^{\hat{\mathcal{P}}}}^{\hat{\mathcal{P}}}(\cdot, \cdot, \cdot, \cdot)$ in (7) based on the resulting optimal partition $\hat{\mathcal{P}}$. In our applications of Section 3, the optimal selected subtrees always implied a maximal number of three thresholds, which is strictly lower than the number of thresholds for the maximal tree in the first step of our estimation procedure. This confirms the absence of over-fitting in all our in-sample results of Section 3.

3. EMPIRICAL RESULTS

This section presents the results of our estimations of double tree-structured AR-GARCH models on returns series of some global stock market indices.

3.1. Data

We consider daily (log) return series of nine major stock indices: the French CAC40 Index, the German Deutsche Aktien Index (DAX30), the Italian BCI General Index, the Canadian Toronto SE35 Index, the UK FT-SE-A All-Share (FTSE100) Index, the Japanese NIKKEI225 Average Index, the Swiss SMI Index, the Hang Seng Index and the US S&P500 Index. Data are for the time period between January 1, 1998 and November 4, 2002, for a total of 1262 trading days, and have been downloaded from *Datastream International*.

We split the sample period into two subperiods and use the first 781 observations (until December 29, 2000) as in-sample data (for estimation purposes) and the remaining 481 observations as out-of-sample data (for forecasting performance evaluation purposes). As mentioned, we fix the maximal number of candidate thresholds (regimes) in our tree-structured models at four (five). This implies a number of estimated parameters across the selected models with US index returns which is typically about 15 and never more than 20. We feel that this is a quite reasonable model dimension, given the sample size of our in-sample period. Moreover, when testing for parameter significance we always apply model-based bootstrapped confidence intervals (see Efron and Tibshirani, 1993) to improve the finite sample accuracy of our interval estimates.

Summary statistics of in-sample daily returns for the above stock indices are presented in Table I. Sample means for the Canadian, Hong Kong and European index returns are similar. The Japanese market shows a negative mean return that is attributable to a bear market during the considered in-sample period; the market performed even worse during the following Asian crisis and the more recent US recession. The sample standard deviation exhibited by the Hang Seng returns is, as expected, considerably higher than that of all other stock index returns. The Ljung–Box statistics LB(10) testing for autocorrelations in the level of returns up to the 10th order show, in most cases, significance, rejecting the hypothesis of the absence of autocorrelation in daily index returns.[12] The |LB(10)| statistics for examining the null hypothesis of dependency of the absolute index returns are all highly significant, supporting a volatility clustering hypothesis. The Canadian

[12] The existence of this autocorrelation may result from some market frictions or some slow market adjustments.

Table I. Summary statistics on index returns of eight stock indices for the time period between January 1, 1998 and December 29, 2000, for a total of 781 in-sample observations. Sample sdev, LB(10) and |LB(10)| are the sample standard deviations and the Ljung–Box statistics testing for autocorrelation in the level of returns and the level of absolute returns, respectively, up to the 10th lag. Asterisks indicate statistical significance at the 5% level. Corr. with US are the sample correlations of the different daily index returns for the eight markets under scrutiny with the daily S&P500 returns

Index	Sample mean	Sample sdev	LB(10)	|LB(10)|	Corr. with US
CAC40	0.0872	1.4354	18.891*	62.689*	0.4318
DAX30	0.0531	1.5672	22.429*	170.31*	0.4154
BCI General	0.0758	1.4654	20.106*	230.44*	0.2983
Toronto SE35	0.0579	1.3148	11.186	128.14*	0.6883
FTSE100	0.0246	1.2035	24.383*	68.351*	0.3994
NIKKEI225	−0.0130	1.4496	7.8315	48.472*	0.0809
SMI	0.0334	1.2885	19.272*	534.37*	0.3771
Hang Seng	0.0438	2.1402	15.550	49.702*	0.1166

stock index returns exhibit the highest sample correlations with US S&P500 returns, whereas the lowest correlations are those of the Asian stock index returns.

3.2. Estimation Results

This section presents our estimated double tree-structured AR-GARCH models. Past lagged returns of the US S&P500 Index are used as conditioning information to predict X_t, i.e. X_t^{us} is in all our estimations the S&P500 return at day t. Estimated parameters and thresholds in our tree-structured AR-GARCH models are summarized in Tables II and III, for (i) the case where lagged US index returns are incorporated in the model, i.e. $\mathbf{X}_t = (X_t, X_t^{\mathrm{us}})'$ and (ii) the case where they are not, i.e. $\mathbf{X}_t = X_t$. More precisely, when it is included in our tree-structured models lagged US market information can affect both the local AR-GARCH structures and the thresholds of the conditional mean and variance functions.[13]

We discuss first the structure of the estimated threshold functions in our model and, in a second step, the parameter estimates obtained for the local AR-GARCH structures.

Estimated Number and Type of Regimes

In Table II we first observe that the estimated threshold functions often involve more than one (multivariate) threshold, i.e. more than two regimes, in conditional means and volatilities of index returns. With the exception of the BCI (where no threshold could be identified) and the FTSE (where only two regimes have been found), the estimated number of regimes is in most of the cases three (the SE35, the NIKKEI, the SMI and the Hang Seng) and in some cases four (the CAC and the DAX). Moreover, our results in Table II show that lagged US index returns do appear in the estimated threshold structure for all models where more than one threshold has been identified. A comparison of Tables II and III shows that the threshold functions estimated when including lagged US index information exhibit typically a richer structure than that of models where no US

[13] For tree-structured models the case where no US market information is included corresponds to estimating a model where the mean equation represents a classical autoregressive term for X_t and where the thresholds are estimated in the (lower dimensional) state space $G = \mathbb{R} \times \mathbb{R}^+$.

Table II. Estimated double tree-structured AR-GARCH models for the returns of eight stock market indices, when incorporating lagged US index returns. Data are for the time period between January 1, 1998 and December 29, 2000, for a total of 781 in-sample observations. Model-based bootstrap standard errors for the coefficients in the regime's mean equation are given in parentheses. * and ** indicate statistical significance at the 5% and 1% level. Reg. av. var is the average variance of the different local AR-GARCH models, measured by $\alpha_{0,j}/(1-\alpha_{1,j}-\beta_j)$, where $j=1,\ldots,\#$ of regimes. The local GARCH parameters that are statistically significant at the 1% level are listed under the corresponding average variance

Index	# Reg.	Regime's structure	Regime's mean eq.	Reg. av. var
CAC40	4	$x \leq 0.508$	$-0.1317x + 0.4193x^{us}$ (0.022)** (0.023)**	0.0432 α_1, β
		$x > 0.508, x^{us} \leq 0, \sigma^2 \leq 1.572$	$0.0406x + 0.5738x^{us}$ (0.020)* (0.038)**	1.1581 α_0, β
		$x > 0.508, x^{us} \leq 0, \sigma^2 > 1.572$	$0.0834x + 0.5878x^{us}$ (0.027)** (0.032)**	3.5378 $\alpha_0, \alpha_1, \beta$
		$x > 0.508$ and $x^{us} > 0$	$-0.0018x + 0.3103x^{us}$ (0.027) (0.034)**	0.3133 β
DAX30	4	$x \leq -0.843$	$-0.1344x + 0.5010x^{us}$ (0.049)** (0.080)**	0.1644 β
		$x > -0.843$ and $x \leq 0.550$	$-0.3247x + 0.4155x^{us}$ (0.181)* (0.068)**	1.8875 β
		$x > 0.550$ and $x^{us} \leq -1.281$	$-0.8798x + 0.0717x^{us}$ (0.221)** (0.195)	0.0004 β
		$x > 0.550$ and $x^{us} > -1.281$	$0.0017x + 0.2656x^{us}$ (0.049) (0.075)**	0.3988 β
BCI General	1	no threshold	$-0.0239x + 0.3574x^{us}$ (0.034) (0.031)**	1.9973 α_1, β
Toronto SE35	3	$x^{us} \leq 0.791$ and $x \leq -0.639$	$0.1507x - 0.1561x^{us}$ (0.067)* (0.082)*	$\approx 10^{-9}$ β
		$x^{us} \leq 0.791$ and $x > -0.639$	$0.0507x + 0.2871x^{us}$ (0.057) (0.039)**	$\approx 10^{-5}$ β
		$x^{us} > 0.791$	$-0.1530x + 0.2957x^{us}$ (0.057)** (0.037)**	0.4161 α_0, β
FTSE100	2	$x \leq 0.339$	$-0.0935x + 0.3780x^{us}$ (0.048)* (0.038)**	0.8147 β
		$x > 0.339$	$-0.0724x + 0.2984x^{us}$ (0.046) (0.042)**	0.0001 β
NIKKEI225	3	$x \leq 0.278$ and $x^{us} \leq -1.281$	$-0.2299x + 0.6564x^{us}$ (0.183) (0.092)**	$\approx 10^{-7}$ β
		$x \leq 0.278$ and $x^{us} > -1.281$	$-0.0106x + 0.3842x^{us}$ (0.049) (0.065)**	2.1926 β
		$x > 0.278$	$-0.0590x + 0.2650x^{us}$ (0.049) (0.055)**	0.0018 β
SMI	3	$x \leq -1.202$	$-0.0846x + 0.3952x^{us}$ (0.070) (0.086)**	0.8798 β
		$x > -1.202$ and $x^{us} \leq 1.327$	$-0.0085x + 0.2288x^{us}$ (0.046) (0.036)**	0.6863 β
		$x > -1.202$ and $x^{us} > 1.327$	$-0.1269x + 0.0974x^{us}$ (0.103) (0.047)*	0.0009 β
Hang Seng	3	$x \leq -1.175$ and $x^{us} \leq -0.669$	$-0.4966x + 1.2376x^{us}$ (0.017)** (0.229)**	$\approx 10^{-5}$ β
		$x \leq -1.175$ and $x^{us} > -0.669$	$0.0969x + 0.8324x^{us}$ (0.069) (0.134)**	$\approx 10^{-8}$ α_1, β
		$x > -1.175$	$0.0796x + 0.6628x^{us}$ (0.042)* (0.052)**	1.7984 β

Table III. Estimated double tree-structured AR-GARCH models for the returns of eight stock market indices, without incorporating lagged US index returns. Data are for the time period between January 1, 1998 and December 29, 2000, for a total of 781 in-sample observations. Model-based bootstrap standard errors for the coefficients in the regime's mean equation are given in parentheses. * and ** indicate statistical significance at the 5% and 1% level. Reg. av. var is the average variance of the different local AR-GARCH models, measured by $\alpha_{0,j}/(1-\alpha_{1,j}-\beta_j)$, where $j = 1, \ldots, \#$ of regimes. The local GARCH parameters that are statistically significant at the 1% level are listed under the corresponding average variance

Index	# Reg.	Regime's structure	Regime's mean eq.	Reg. av. var
CAC40	2	$x \leq 0.508$	$0.0256x$ (0.058)	1.9876 α_1, β
		$x > 0.508$	$0.1208x$ (0.046)**	0.0321 β
DAX30	2	$x \leq 0.550$	$0.0162x$ (0.055)	0.7522 α_1, β
		$x > 0.550$	$0.0681x$ (0.051)	0.2861 α_1, β
BCI General	2	$x \leq 0.112$	$0.0564x$ (0.051)	1.8925 α_1, β
		$x > 0.112$	$0.0983x$ (0.054)*	0.0001 α_1, β
Toronto SE35	2	$x \leq 0.463$	$0.0236x$ (0.049)	0.1471 α_1, β
		$x > 0.463$	$0.1199x$ (0.046)**	0.3183 β
FTSE100	3	$x \leq -0.264$	$0.0801x$ (0.055)	0.1128 β
		$-0.264 < x \leq 0.015$	$0.9430x$ (0.713)	1.2816 β
		$x > 0.015$	$0.0735x$ (0.047)	$\approx 10^{-5}$ β
NIKKEI225	3	$x \leq -1.542$	$0.0097x$ (0.057)	0.0004 β
		$-1.542 < x \leq 0.278$	$0.1362x$ (0.102)	2.4193 β
		$x > 0.278$	$-0.0476x$ (0.054)	0.0290 β
SMI	3	$x \leq 0.026$	$0.0372x$ (0.059)	0.4498 α_1, β
		$0.026 < x \leq 1.377$	$-0.0723x$ (0.074)	0.0096 β
		$x > 1.377$	$0.0785x$ (0.056)	0.0002 β
Hang Seng	2	$x \leq -2.064$	$0.0328x$ (0.153)	0.0001 β
		$x > -2.064$	$0.1096x$ (0.177)	2.1733 β

market information is incorporated. Exceptions to this rule are the FTSE index (where in the first case two regimes and in the second case three regimes have been identified) and the BCI index (where in the first case one regime and in the second case two regimes have been identified).

Second, the structure of estimated thresholds in conditional means and volatilities is associated with a different impact of 'good' and 'bad' lagged domestic index returns. For instance, Table II shows that a first threshold separating 'bad' and 'good' lagged domestic index returns has been found for the CAC, the DAX, the FTSE, the NIKKEI and the Hang Seng indices. For the BCI index no threshold has been identified. For the Canadian Toronto SE35 Index, instead, a threshold determined by 'bad' and 'good' lagged US market returns is the main determinant of asymmetries in conditional means and volatilities. For this latter case, the negligible impact of domestic index returns is possibly due to the well-known strong interaction of the US and Canadian markets and their higher synchronicity. Table I highlights the clearly higher correlation of US and Canadian index returns, compared to the other markets analysed.

Third, lagged US index returns do affect the structure of the thresholds in the estimated models. Typically, they induce two further subregimes in response either to 'good' or 'bad' lagged domestic returns, depending on the geographic location of the market. For European indices like the CAC, the DAX and the SMI it appears that 'good' lagged domestic returns imply a different local AR-GARCH regime under either 'bad' or 'good' lagged US index returns. Such indices are characterized by one regime in response to 'bad' domestic index returns and at least two further regimes in response to (i) 'good' domestic and 'bad' US index returns or (ii) 'good' domestic and US index returns. Only for the CAC lagged domestic volatility plays a role in the estimated tree-structured AR-GARCH structure.[14] Therefore, lagged information from the US market seems to play an important role in forecasting domestic index returns precisely when the investors' perception of the market based on domestic conditioning information is not too negative. When focusing on the indices from the Asian area, it appears that the role of lagged US index returns in the estimated mean and volatility thresholds is different. For the NIKKEI and the Hang Seng, S&P500 lagged returns determine different regimes in conditional means and volatilities when lagged domestic index returns are below a given threshold. In these states, lagged US market information plays an important role to forecast domestic index returns. This difference with some of the results for European indices is possibly due to the bearish market that prevailed over our in-sample period in the Asian area.

Estimated Local Conditional Mean and Variance Functions

When analysing in more detail the estimated local mean and variance functions of our tree-structured AR-GARCH models, some further features arise. In Tables II and III we observe that the estimated thresholds vary considerably across markets. Therefore, a model of the index returns dynamics based on mean and volatility thresholds only at 0 (as in several models in the literature) seems not to be supported by our in-sample analysis. The local parameter estimates of the mean equation in Table II show that the loading parameter ψ_j for lagged S&P500 returns is essentially always positive and typically much larger than the loading parameter ϕ_j for lagged domestic returns. This finding supports the hypothesis that the impact of lagged US index returns is larger and more systematic than that of lagged domestic index returns. Moreover, it is consistent with the evidence that US market information is rapidly transmitted to the rest of the world and that the US market provides price leadership in the equity world market (see for instance Eun and Shin, 1989; Chiang, 1998; Masih and Masih, 2001). These findings are also naturally compatible with the empirical fact that international investors often over-react to US index returns shocks, while they are less sensitive to other markets (Becker *et al.*, 1995).

[14] Conditional means and volatilities are characterized in this case by four regimes. Two of them are induced by a threshold on past domestic volatilities; see again Table II.

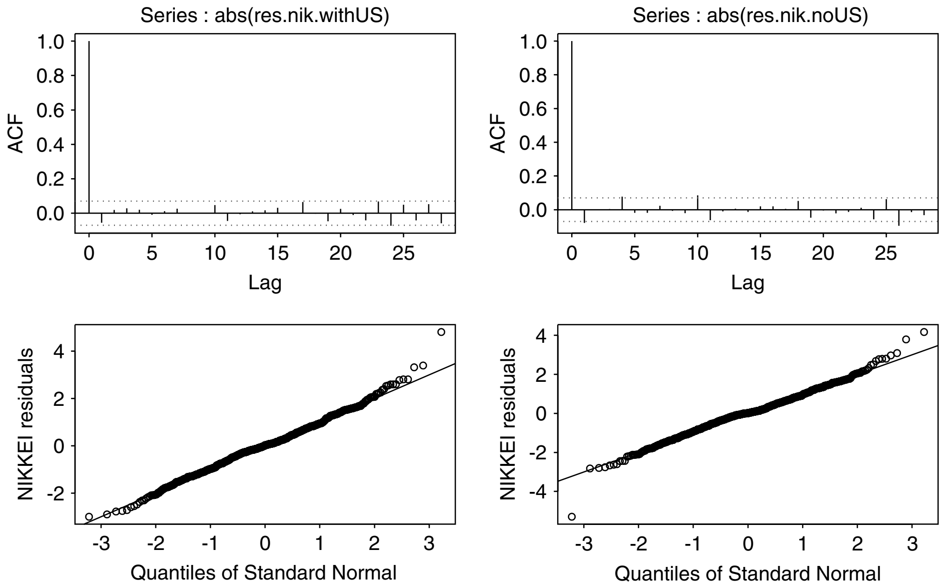

Figure 2. Results for returns of the NIKKEI225 index based on a double tree-structured AR-GARCH model fit incorporating US news (left panel) and not incorporating US news (right panel). Data are for the time period between January 1, 1998 and December 29, 2000, for a total of 781 in-sample observations. Top and bottom panels present the estimated autocorrelation functions of absolute residuals $|\hat{Z}_t|$ and the normal plots of the residuals $\hat{Z}_t$, respectively

Finally, the local variance structures estimated in our model also imply quite different volatility levels and dependencies across different regimes. For instance, when comparing the local regime average variances in Table II, as measured by $\alpha_{0,j}/(1-\alpha_{1,j}-\beta_j)$, we observe that the asymmetric patterns of conditional volatilities generate a quite broad spectrum of local volatility dynamics and averages (see, for instance, the corresponding column for the CAC in Table II).

To illustrate some basic graphical diagnostic checks on our results, Figure 2 presents some more details on the NIKKEI residuals obtained from the estimated double tree-structured AR-GARCH model for this market, both when including and excluding conditioning lagged information US index returns. Similar findings arise for the rest of the estimated tree-structured models.

Graphical diagnostics for the residuals are satisfactory for both tree-structured models, showing a slight tendency towards heavier tails than under a standard normal error distribution. The Ljung–Box statistics testing for autocorrelation in the resulting absolute residuals are 6.1368 and 16.6257, respectively (when including and not including foreign news, respectively), in both cases not significant at the 5% level.

3.3. In- and Out-of-Sample Performance Results

We evaluate in our application the in-sample and out-of-sample performance of tree-structured AR-GARCH models relative to a standard AR-GARCH model and a GJR AR-GARCH model incorporating an asymmetric impact of lagged domestic returns shocks in the volatility dynamics. These are short memory models. We therefore also estimate an AR-FIGARCH(1,d,1) model

(Baillie *et al.*, 1996) to compare the forecasting ability of tree-structured models with a benchmark model that explicitly allows for long memory in second moments.

We estimate the parameters of the AR-FIGARCH(1,d,1) model based on a longer in-sample period (from January 1, 1990 to December 29, 2000, consisting of 2869 observations) than the one used for the other models, in order to allow for a more precise estimation of a possible long memory pattern in second moments of index returns. We also investigate the in-sample and out-of-sample performance of AR-FIGARCH(1,d,1) models using the shorter in-sample period from January 1, 1998 to December 29, 2000. However, the out-of-sample results were qualitatively very similar, although slightly less favourable. Therefore, in the FIGARCH case we report results based only on the longer in-sample period.

We make a distinction between models where lagged US market information is incorporated and models where it is not. For the AR-GARCH, the GJR AR-GARCH and the AR-FIGARCH models lagged US returns are included by means of a further regressor in the linear autoregressive mean equation. For the tree-structured models, lagged US market information can affect both the local AR-GARCH structures and the thresholds of the conditional mean and variance functions.

The in- and out-of-sample performance of the different models is quantified by means of several performance measures. In-sample performance is measured by means of the following statistics:

- The AIC statistic in (11).
- The in-sample prediction loss

$$\text{IS-PL}_2 = n^{-1}\sum_{t=1}^{n}(\widehat{\sigma}_t^2 - (X_t - \widehat{\mu}_t)^2)^2$$

- The in-sample heteroskedastic-adjusted mean squared error (Bollerslev and Ghysels, 1996)

$$\text{IS-HMSE} = n^{-1}\sum_{t=1}^{n}\left(\frac{(X_t - \widehat{\mu}_t)^2}{\widehat{\sigma}_t^2} - 1\right)^2$$

where $\widehat{\theta}$, $\widehat{\mu}_t$ and $\widehat{\sigma}_t^2$ are based on estimates from the in-sample data $\mathbf{X}_1^n$, for a sample size $n = 781$.

Performance measures based on the loss function implied by IS-PL and IS-HMSE are more relevant when we compare out-of-sample performance. In our analysis, out-of-sample performance is measured by means of the following statistics:

- The out-of-sample negative pseudo log-likelihood

$$\text{OS-NL} = -\ell(\hat{\theta}; \mathbf{Y}_2^{n_{\text{out}}})$$

- The out-of-sample prediction loss

$$\text{OS-PL}_2 = n_{\text{out}}^{-1}\sum_{t=1}^{n_{\text{out}}}(\widehat{\sigma}_t^2(\mathbf{Y}_1^{t-1}) - (Y_t - \widehat{\mu}_t(\mathbf{Y}_1^{t-1}))^2)^2$$

- The out-of-sample heteroskedastic-adjusted mean squared error

$$\text{OS-HMSE} = n_{\text{out}}^{-1}\sum_{t=1}^{n_{\text{out}}}\left(\frac{(Y_t - \widehat{\mu}_t(\mathbf{Y}_1^{t-1}))^2}{\widehat{\sigma}_t^2(\mathbf{Y}_1^{t-1})} - 1\right)^2$$

where $t = 1, \ldots, n_{\text{out}}$ is an index that now indexes out-of-sample observations $\mathbf{Y}_t = (Y_t, Y_t^{\text{us}})$ of domestic and foreign index returns and $\mathbf{Y}_1^t = (\mathbf{Y}_1, \ldots, \mathbf{Y}_t)$. $\widehat{\mu}_t(\mathbf{Y}_1^{t-1})$ and $\widehat{\sigma}_t^2(\mathbf{Y}_1^{t-1})$ are the conditional mean and volatility functions estimated from in-sample data and evaluated on out-of-sample data. The sample size of our out-of-sample exercise is $n_{\text{out}} = 481$.

Table IV summarizes the in- and out-of-sample performance measures estimated for a standard AR(1)-GARCH(1,1) model, a GJR-AR(1)-GARCH(1,1) model, an AR(1)-FIGARCH(1,d,1) model and a tree-structured AR(1)-GARCH(1,1) model in our application. For these models we compute the above performance measures when (i) lagged US market returns are included in the model (the first four rows for each market in Table IV) and (ii) they are not (the second four rows for each market in Table IV). As mentioned, for the AR-GARCH, the GJR AR-GARCH and the AR-FIGARCH settings lagged US returns are included in the models by means of a further regressor in the linear autoregressive mean equation.

From Table IV we observe that tree-structured models yield substantial improvements of all performance statistics, both when incorporating and when excluding lagged US market information in the model. Average gains in the OS-NL, the OS-PL$_2$ and the OS-HMSE measures over a standard AR-GARCH fit are about 1%, 4% and 30%, without incorporating lagged US market information, and about 1%, 3% and 35%, when including S&P500 lagged index returns. The average gains of tree-structured models over a GJR-AR-GARCH or an AR-FIGARCH model fit are smaller, but still important for prediction purposes, as we will highlight in more detail in the next section. It is interesting to note that even if the largest gains in out-of-sample goodness of fit measures over simple AR-GARCH(1,1) models with US news are obtained for CAC and DAX index returns (1.5%, 6% and 54% on average for the OS-NL, the OS-PL$_2$ and the OS-HMSE measures), which are fitted by models with the largest number of regimes, the resulting performance gains are not necessarily directly related to the estimated number of regimes. For example, the average gain in the OS-NL measure for tree-structured models with three regimes (as for instance for the Toronto SE35, the NIKKEI, the SMI and the Hang Seng) is smaller (0.5% vs. 0.7%) than that resulting from the analysis of the FTSE (which is fitted by a model with two regimes).

Lagged US market returns have strong predictive potential when they are included in the thresholds of our tree-structured models: tree-structured models incorporating lagged US S&P returns yield substantial improvements of most performance measures. For example, the average gains in the OS-NL and the OS-PL$_2$ of double tree-structured AR-GARCH models including lagged US index returns are about 2% and 8%, compared to a model without US market lagged information. The results obtained when comparing tree-structured models with and without US news based on the OS-HMSE performance measure are not uniform. The same holds for the GJR-AR-GARCH and the AR-FIGARCH models. A formal statistical test for the additional predictive power of tree-structured models including lagged US market information is performed in Section 3.5.

3.4. Testing Differences in Forecasting Ability

In this subsection we test formally whether differences in the out-of-sample model performances, as highlighted in the last section, are statistically significant. We consider models which do not include lagged US index returns. The additional value of including US index returns is tested explicitly in the next section. We make use of the t-type and the sign-type performance tests, as proposed by Audrino and Bühlmann (2004), extending previous work by Diebold and Mariano

Table IV. Goodness of fit measures for the returns of eight stock market indices when estimating a classical AR(1)-GARCH(1,1) model (simple no US), a standard GJR-AR(1)-GARCH(1,1) model (GJR no US), a standard AR(1)-FIGARCH(1,d,1) model (FIGARCH no US), a standard double tree-structured AR-GARCH model without lagged US index returns (tree no US), an AR-GARCH(1,1)-type model including lagged US index returns in the mean equation (simple with US), a GJR-AR-GARCH(1,1)-type model including lagged US index returns in the mean equation (GJR with US), an AR-FIGARCH(1,d,1)-type model including lagged US index returns in the mean equation (FIGARCH with US) and a double tree-structured AR-GARCH model incorporating lagged US index returns as described in Section 2.3 (tree with US). In-sample data are for the time period between January 1, 1998 and December 29, 2000 (for a total of 781 in-sample observations) for all models except the FIGARCH models, where we consider a longer in-sample estimation period beginning in January 1, 1990 (for a total of 2869 observations). However, for comparison purposes the in-sample statistics of the FIGARCH models are reported for the common in-sample period (except for the AIC statistic). The out-of-sample test data are the 481 observations from December 30, 2000 until November 4, 2002. The different performance measures listed in the table have been defined in Section 3.3

Index	Model	AIC	OS-NL	IS-PL$_2$	OS-PL$_2$	IS-HMSE	OS-HMSE
CAC40	simple no US	2721.8	946.51	11.862	41.380	2.3455	3.7576
	GJR no US	2716.2	938.51	11.533	40.755	2.4265	2.8365
	FIGARCH no US	—	946.82	11.975	41.681	2.1697	3.4430
	tree no US	2708.6	934.01	11.352	39.607	2.3643	2.2847
	simple with US	2644.6	919.78	10.247	35.404	2.4298	4.4444
	GJR with US	2636.6	913.04	9.8791	34.902	2.4450	3.1437
	FIGARCH with US	—	923.69	10.301	36.165	2.2919	4.2737
	tree with US	2634.0	910.19	9.3619	34.487	2.3533	2.2122
DAX30	simple no US	2834.7	983.98	16.930	60.407	2.6412	2.8587
	GJR no US	2829.3	973.34	16.644	58.329	2.5394	2.1909
	FIGARCH no US	—	976.63	17.261	58.586	2.6679	2.3654
	tree no US	2817.9	974.76	16.733	57.233	1.8743	1.8851
	simple with US	2758.7	980.91	13.275	60.641	2.6412	3.5575
	GJR with US	2753.3	971.33	13.090	58.955	2.6303	2.7203
	FIGARCH with US	—	983.96	12.601	61.276	3.3110	3.2802
	tree with US	2741.7	961.07	12.725	54.973	2.6252	1.4525
BCI General	simple no US	2666.9	838.87	15.205	20.762	2.6399	3.4227
	GJR no US	2668.8	837.67	15.214	20.653	2.6611	3.2706
	FIGARCH no US	—	846.37	15.621	21.686	2.1193	2.9338
	tree no US	2666.5	836.64	15.525	20.527	2.7703	2.7257
	simple with US	2582.8	833.39	10.606	20.033	2.4162	3.7488
	GJR with US	2584.7	833.69	10.608	19.983	2.4260	3.6453
	FIGARCH with US	—	848.35	10.602	22.709	2.0201	3.2593
	tree with US	2582.8	833.39	10.606	20.033	2.4162	3.7488
Toronto SE35	simple no US	2557.8	760.57	12.784	7.0193	3.3163	3.1557
	GJR no US	2533.0	753.50	12.518	6.6747	2.8312	2.6079
	FIGARCH no US	—	765.05	13.004	6.9933	4.3227	3.5441
	tree no US	2528.3	759.42	12.505	6.6588	2.7713	1.9665
	simple with US	2548.4	755.63	12.275	6.8458	3.2204	3.3386
	GJR with US	2523.4	749.47	12.019	6.5652	2.7482	2.8557
	FIGARCH with US	—	762.47	12.627	6.8402	3.1924	2.9168
	tree with US	2500.2	750.52	11.398	6.6953	2.6378	1.8603
FTSE100	simple no US	2454.7	828.35	4.8219	18.358	2.3532	3.2353
	GJR no US	2446.8	820.89	4.7218	17.701	2.4545	2.4461
	FIGARCH no US	—	820.24	4.8575	17.611	2.8134	2.7872
	tree no US	2443.6	815.50	4.5815	17.145	2.2795	1.8845

(*continued overleaf*)

Table IV. (*Continued*)

Index	Model	AIC	OS-NL	IS-PL$_2$	OS-PL$_2$	IS-HMSE	OS-HMSE
	simple with US	2374.2	804.27	3.9108	14.291	2.2782	3.9161
	GJR with US	2361.3	798.89	3.7808	13.899	2.3056	3.0167
	FIGARCH with US	—	796.13	3.9520	13.965	2.8675	3.2294
	tree with US	2363.5	798.20	3.7656	14.064	2.3564	2.4177
NIKKEI225	simple no US	2767.2	934.27	17.525	26.539	3.9817	3.9232
	GJR no US	2762.4	933.30	17.514	26.532	4.0245	3.6771
	FIGARCH no US	—	932.20	17.868	26.414	3.7295	3.0863
	tree no US	2755.0	932.42	16.739	26.517	3.5154	3.7015
	simple with US	2656.9	912.05	12.413	25.655	3.4317	5.1260
	GJR with US	2654.4	909.42	12.386	25.506	3.4674	4.9817
	FIGARCH with US	—	910.61	12.5959	25.629	2.8834	4.0795
	tree with US	2643.3	910.86	11.008	23.777	2.9058	3.7660
SMI	simple no US	2417.8	829.38	11.546	31.109	3.0787	4.5177
	GJR no US	2399.8	822.41	10.722	29.368	3.0315	3.0899
	FIGARCH no US	—	821.98	11.960	31.005	2.8716	4.4005
	tree no US	2396.5	819.93	10.341	29.179	2.6104	2.8377
	simple with US	2387.7	814.54	11.022	27.405	2.9559	4.8466
	GJR with US	2368.0	809.28	10.213	26.186	2.8714	3.5886
	FIGARCH with US	—	808.08	11.702	26.816	2.9464	4.5932
	tree with US	2362.6	802.83	10.218	28.383	2.5937	3.1747
Hang Seng	simple no US	3335.5	870.13	111.03	25.746	3.7078	2.6759
	GJR no US	3313.1	869.06	106.97	25.925	3.4982	1.8692
	FIGARCH no US	—	868.05	111.21	25.613	3.6688	2.5597
	tree no US	3306.9	868.18	104.80	25.677	3.4791	1.6807
	simple with US	3174.9	840.91	93.215	23.772	3.2626	3.9421
	GJR with US	3155.6	843.81	90.009	23.819	3.0472	3.2645
	FIGARCH with US	—	831.45	93.856	23.203	3.2860	3.8145
	tree with US	3144.2	836.45	85.962	22.823	2.5982	3.1156

(1995). We test for significance of the difference in the OS-NL, the OS-PL$_2$ and the OS-HMSE performance measures of tree-structured AR-GARCH models against the GJR-AR(1)-GARCH(1,1) and the AR(1)-FIGARCH(1,d,1) models.

The tests are defined as follows. Let $\widetilde{U}_t$ be the realized out-of-sample loss associated at time t with a given model and based on a given loss function $\widetilde{U}$. By applying a suitable functional form for $\widetilde{U}$ we have, for instance:

$$\sum_{t=1}^{n_{\text{out}}} \widetilde{U}_{t;\text{model}} = \text{OS-NL}, \quad \sum_{t=1}^{n_{\text{out}}} \widetilde{U}_{t;\text{model}} = \text{OS-PL}_2 \quad \text{or} \quad \sum_{t=1}^{n_{\text{out}}} \widetilde{U}_{t;\text{model}} = \text{OS-HMSE}$$

The realized loss difference at time t between Model$_1$ and Model$_2$ is

$$\widehat{D}_t = \widetilde{U}_{t;\text{Model}_1} - \widetilde{U}_{t;\text{Model}_2}, \quad t = 1, \ldots, n_{\text{out}}$$

We test the null hypothesis that the differences $\widehat{D}_t$ have mean zero against the alternative of mean less than zero, i.e. the hypothesis that the average losses implied by Model$_1$ are smaller than those of Model$_2$. In all our empirical tests Model$_1$ will be a tree-structured model while Model$_2$ will be any of the competing volatility models. Thus, a negative value of the t-type statistic in our application is associated with a higher forecasting power of tree-structured models. The t-type test statistic is

$$\sqrt{n_{\text{out}}}\frac{\overline{D}}{\widehat{\sigma}_{D;\infty}}, \quad \text{where } \overline{D} = \frac{1}{n_{\text{out}}}\sum_{t=1}^{n_{\text{out}}}\widehat{D}_t \tag{12}$$

with $\widehat{\sigma}^2_{D;\infty} = (2\pi)\widehat{f}_{\widehat{D}}(0)$ and $\widehat{f}_{\widehat{D}}(0)$ a smoothed periodogram estimate at frequency zero based on $\widehat{D}_1, \ldots, \widehat{D}_{n_{\text{out}}}$; see for example Brockwell and Davis (1991). Then, under the given null hypothesis

$$\sqrt{n_{\text{out}}}\frac{\overline{D}}{\widehat{\sigma}_{D;\infty}} \Longrightarrow \mathcal{N}(0, 1) \quad (n_{\text{out}} \longrightarrow \infty) \tag{13}$$

where

$$\sigma^2_{D;\infty} = \sum_{k=-\infty}^{+\infty} \text{Cov}[\widehat{D}_0, \widehat{D}_k] = 2\pi f_{\widehat{D}}(0) \tag{14}$$

and $f_{\widehat{D}}(0)$ is the spectral density of $\{\widehat{D}_t\}_t$ at zero. The sign-type test[15] is based on the sequence of Bernoulli random variables

$$\widehat{W}_t = I_{\{\widehat{D}_t \leq 0\}}, \quad t = 1, \ldots, n_{\text{out}}$$

that indicate the event $\{\widehat{D}_t \leq 0\}$, i.e. the event that the loss difference between Model_1 and Model_2 is negative (see again the above description for the t-type test). This test is devoted to test the null hypothesis that the mean number of negative differences $\widehat{D}_t$ is $\frac{1}{2}$ against the alternative that it is larger than $\frac{1}{2}$. The sign-type test has a better signal to noise ratio than the t-type test (Audrino and Bühlmann, 2004) and is more robust in the presence of outliers and aberrant observations. The sign-type test statistic is given by

$$\sqrt{n_{\text{out}}}\frac{\overline{W} - \frac{1}{2}}{\widehat{\sigma}_{W;\infty}}, \quad \text{where } \overline{W} = \frac{1}{n_{\text{out}}}\sum_{t=1}^{n_{\text{out}}}\widehat{W}_t \tag{15}$$

where $\widehat{\sigma}^2_{W;\infty}$ is as in (12) but based on $\widehat{W}_1, \ldots, \widehat{W}_{n_{\text{out}}}$. Positive values of the sign-type statistic indicated a higher forecasting power of Model_1 relative to a competing Model_2. As for the t-type test, also for our sign-type tests in the application part Model_1 will always be a tree-structured model while Model_2 will be any of the competing volatility models. Similarly to the above t-type test, we have

$$\sqrt{n_{\text{out}}}\frac{\overline{W} - \frac{1}{2}}{\widehat{\sigma}_{W;\infty}} \Longrightarrow \mathcal{N}(0, 1) \quad (n_{\text{out}} \longrightarrow \infty) \tag{16}$$

under the given null hypothesis. The results of the t-type and sign-type tests for the eight equity indices under scrutiny are presented in Table V for the loss functions $\widetilde{U}$ implied by the OS-PL_2, the OS-NL and the OS-HMSE performance indicators.

In the upper part of Table V we consider differences in forecasting performance of the tree-structured AR-GARCH models against a GJR-AR(1)-GARCH(1,1) model. In four out of eight

[15] See also the related work by Pesaran and Timmermann (1992).

Table V. Tests for a difference in the out-of-sample performance of double tree-structured AR-GARCH models against standard GJR-AR-GARCH models (top) and against AR-FIGARCH models (bottom). The table gives the values of the relevant test statistics and the corresponding p-values (below, in parentheses). Columns 2, 4 and 6 present generalized t-type tests on the series of differences of performance losses $\tilde{U}_t$. Columns 3, 5 and 7 present generalized sign-type tests on the series of differences of performance losses $\tilde{U}_t$

Tree against GJR

Index	Performance measure					
	OS-PL_2		OS-NL		OS-HMSE	
	t-type	sign-type	t-type	sign-type	t-type	sign-type
CAC40	−0.934	−0.454	−0.591	−0.172	−1.626	1.657
	(0.175)	(0.325)	(0.277)	(0.432)	(0.052)	(0.048)
DAX30	−0.752	−0.358	0.149	−0.350	−1.448	−0.019
	(0.226)	(0.360)	(0.441)	(0.363)	(0.074)	(0.492)
BCI General	−0.302	−0.711	−0.255	−0.685	−2.418	−0.706
	(0.382)	(0.239)	(0.399)	(0.247)	(0.008)	(0.240)
Toronto SE35	−0.049	−2.109	0.957	−1.499	−1.673	−1.408
	(0.481)	(0.017)	(0.169)	(0.067)	(0.047)	(0.080)
FTSE100	−1.791	−0.062	−0.791	−0.884	−1.069	0.849
	(0.037)	(0.475)	(0.214)	(0.188)	(0.143)	(0.198)
NIKKEI225	−0.092	−0.789	−0.179	−1.165	−0.029	0.694
	(0.463)	(0.215)	(0.429)	(0.122)	(0.489)	(0.244)
SMI	−0.250	−0.082	−0.086	−0.539	−0.324	−0.378
	(0.401)	(0.467)	(0.466)	(0.295)	(0.373)	(0.353)
Hang Seng	−0.328	0.265	−0.208	−0.014	−0.375	2.028
	(0.372)	(0.394)	(0.417)	(0.494)	(0.354)	(0.021)

Tree against FIGARCH

Index	Performance measure					
	OS-PL_2		OS-NL		OS-HMSE	
	t-type	sign-type	t-type	sign-type	t-type	sign-type
CAC40	−1.681	0.373	−1.668	0.255	−1.016	1.640
	(0.046)	(0.355)	(0.048)	(0.399)	(0.155)	(0.051)
DAX30	−0.610	−0.169	−0.307	−0.189	−1.773	0.120
	(0.271)	(0.433)	(0.379)	(0.425)	(0.038)	(0.452)
BCI General	−0.530	0.089	−2.286	0.669	−0.411	−0.029
	(0.298)	(0.465)	(0.011)	(0.252)	(0.341)	(0.488)
Toronto SE35	−0.567	−1.431	−0.473	−1.285	−2.031	−1.002
	(0.285)	(0.076)	(0.318)	(0.099)	(0.021)	(0.158)
FTSE100	−1.261	−1.632	−0.597	−1.313	−2.677	−0.554
	(0.104)	(0.051)	(0.275)	(0.095)	(0.004)	(0.290)
NIKKEI225	0.181	0.015	0.050	0.178	0.399	1.417
	(0.428)	(0.494)	(0.480)	(0.429)	(0.345)	(0.078)
SMI	−0.959	0.028	−0.710	−0.335	−1.650	−0.251
	(0.169)	(0.489)	(0.239)	(0.369)	(0.049)	(0.401)
Hang Seng	−0.363	−3.528	0.069	−4.311	−0.958	−3.431
	(0.359)	(0.0002)	(0.497)	(10^{-6})	(0.169)	(0.0003)

markets statistically significant evidence (at the 5% level) in favour of tree-structured models emerges, either based on the OS-NL, the OS-PL_2 or the OS-HMSE t-type and sign-type tests. By

Table VI. Tests for a difference in the out-of-sample performance of double tree-structured AR-GARCH models including lagged US index returns vs. tree-structured models without including lagged US index returns. The table gives the values of the relevant test statistics and the corresponding p-values (below, in parentheses). Columns 2, 4 and 6 present generalized t-type tests on the series of differences of performance losses $\tilde{U}_t$. Columns 3, 5 and 7 present generalized sign-type tests on the series of differences of performance losses $\tilde{U}_t$

Tree with US market returns against Tree no US market returns

Index	Performance measure					
	OS-PL$_2$		OS-NL		OS-HMSE	
	t-type	sign-type	t-type	sign-type	t-type	sign-type
CAC40	−0.724	2.401	−2.889	1.014	−0.228	1.034
	(0.234)	(0.008)	(0.002)	(0.155)	(0.409)	(0.150)
DAX30	−0.264	0.294	−1.302	0.343	−1.029	1.829
	(0.396)	(0.384)	(0.096)	(0.366)	(0.152)	(0.034)
BCI General	−0.167	1.137	−0.238	0.807	0.872	1.608
	(0.434)	(0.128)	(0.406)	(0.210)	(0.192)	(0.054)
Toronto SE35	0.063	0.035	−0.008	0.375	−0.365	0.252
	(0.475)	(0.486)	(0.497)	(0.354)	(0.357)	(0.401)
FTSE100	−0.721	2.715	−2.686	1.030	0.816	1.388
	(0.235)	(0.003)	(0.004)	(0.152)	(0.207)	(0.082)
NIKKEI225	−0.389	2.305	−1.924	2.055	0.051	0.885
	(0.349)	(0.011)	(0.027)	(0.020)	(0.480)	(0.188)
SMI	−0.238	0.968	−1.785	1.058	0.321	0.571
	(0.406)	(0.167)	(0.037)	(0.145)	(0.374)	(0.284)
Hang Seng	−0.226	3.175	−1.786	3.406	0.796	1.119
	(0.411)	(0.0007)	(0.037)	(0.0003)	(0.213)	(0.131)

contrast, only in the case of the Toronto SE35 returns a GJR-AR(1)-GARCH(1,1) fit improves in a statistically significant way upon tree-structured AR-GARCH models. However, in this latter case the test results are not uniform across the different performance measures: for example, the tree-structured model is preferred based on the t-type test based on the OS-HMSE performance measure. When considering statistical evidence at the 10% confidence level, we observe that tree-structured models are significantly better in five out of eight markets. In the lower part of Table V we consider differences of forecasting power of the tree-structured AR-GARCH model against the AR(1)-FIGARCH(1,d,1) model. Again, in six out of eight markets statistically significant evidence (at the 5% level) emerges for a higher predictive power of tree-structured models, either based on the OS-NL, the OS-PL$_2$ or the OS-HMSE t-type and sign-type tests. The tree-structured model is clearly beaten by a FIGARCH fit only in the case of the Hang Seng index returns.

3.5. Testing the Predictive Power of Lagged US Index Returns

We conclude our analysis by testing formally the predictive power of lagged US stock index returns when they are included in the thresholds of tree-structured models. To this purpose, we make use of the t-type and the sign-type tests introduced in the last section. We test for significance of the difference in the OS-NL, the OS-PL$_2$ and OS-HMSE performance measures of tree-structured models including lagged US index returns, relative to a tree-structured model which does not include lagged foreign market information. Results for the t-type and sign-type tests are presented in Table VI, based on the OS-PL$_2$, the OS-NL and the OS-HMSE loss functions $\widetilde{U}$.

Table VI shows that in six out of eight equity markets statistically significant evidence (at the 5% level) for a higher predictive power of tree-structured models incorporating US index returns information emerges, either based on the OS-NL, the OS-PL_2 or the OS-HMSE t-type and sign-type tests. For the Toronto SE35 Index and the BCI General Index the differences in the out-of-sample performances of the competing models are not significant.

4. CONCLUDING REMARKS

We proposed a double tree-structured AR-GARCH model incorporating lagged US index returns to estimate and forecast the volatility dynamics of global stock markets. The model is quite flexible but still parsimonious enough to be statistically and computationally manageable. It can be used to identify and estimate local AR-GARCH structures across several multivariate regimes. We propose a computationally feasible algorithm that can be applied to estimate the model in practice.

In our empirical investigation for eight major stock indices, we produced empirical evidence on the higher predictive potential of tree-structured AR-GARCH models. We found strong evidence in favour of more than two regimes in conditional means and variances of equity index returns. Further, conditioning information from the US market often affects the estimated thresholds and has out-of-sample predictive power, improving the forecasts relative to some competing models in the literature.

Our double tree-structured AR-GARCH methodology is very general and can easily be extended to test for multiple regimes in related applications, where for instance more than one exogenous variable can appear to impact the threshold definitions. Some future research on tree-structured AR-GARCH models covers the application to forecasting implied volatilities and the estimation of general volatility asymmetries and dependencies for individual stock returns.

APPENDIX: ALGORITHM—GROWING THE BINARY TREE

1. Compute the negative log-likelihood (8) based on the trivial partition $\mathcal{P}_{opt}^{(0)} = G$ with

$$g_{\theta^{(0)}}^{\mathcal{P}_{opt}^{(0)}}(x, x^{\mathrm{us}}, \sigma^2) = \phi x + \psi x^{\mathrm{us}}$$

$$f_{\theta^{(0)}}^{\mathcal{P}_{opt}^{(0)}}(\varepsilon, x, x^{\mathrm{us}}, \sigma^2) = \alpha_0 + \alpha_1 \varepsilon^2 + \beta\sigma^2, \theta^{(0)} = (\phi, \psi, \alpha_0, \alpha_1, \beta)$$

   Compute the pseudo maximum likelihood estimate $\hat{\theta}^{(0)}$ of $\theta^{(0)}$ using a quasi-Newton method, cf. Nocedal and Wright (1999). Set $m = 0$.
2. Increment m by one. Search for the best refined partition $\mathcal{P}_{opt}^{(m)}$ under binary splitting of a cell from $\mathcal{P}_{opt}^{(m-1)}$ as in (9). Selection of the optimal partitioning threshold d and coordinate index ι is based on a comparison of the implied reductions in the negative log-likelihood (8). Details on this step are as follows:

(I) Given $\mathcal{P}_{opt}^{(m-1)} = \{\mathcal{R}_1, \ldots, \mathcal{R}_m\}$, consider a new partition $\mathcal{P}^{(m)}$, where a partition cell $\mathcal{R}_{j^*} \in \mathcal{P}^{(m-1)}$ has been split as $\mathcal{R}_{j^*} = \mathcal{R}_{j^*,left} \cup \mathcal{R}_{j^*,right}$. The conditional mean and variance functions associated with the new partition $\mathcal{P}^{(m)}$ are given by

$$g^{\mathcal{P}^{(m)}}_{(\theta^{(m-1)\backslash *},\theta^*)}(x, x^{\mathrm{us}}, \sigma^2) = \sum_{j\neq j^*}(\phi_j x + \psi_j x^{\mathrm{us}})I_{[(x,x^{\mathrm{us}},\sigma^2)\in\mathcal{R}_j]} + \sum_{i\in\{j^*_{left}, j^*_{right}\}}(\phi^*_i x + \psi^*_i x^{\mathrm{us}})I_{[(x,x^{\mathrm{us}},\sigma^2)\in\mathcal{R}_i]} \tag{A.1}$$

and

$$f^{\mathcal{P}^{(m)}}_{(\theta^{(m-1)\backslash *},\theta^*)}(\varepsilon, x, x^{\mathrm{us}}, \sigma^2) = \sum_{j\neq j^*}(\alpha_{0,j} + \alpha_{1,j}\varepsilon^2 + \beta_j\sigma^2)I_{[(x,x^{\mathrm{us}},\sigma^2)\in\mathcal{R}_j]} + \sum_{i\in\{j^*_{left}, j^*_{right}\}}(\alpha^*_{0,i} + \alpha^*_{1,i}\varepsilon^2 + \beta^*_i\sigma^2)I_{[(x,x^{\mathrm{us}},\sigma^2)\in\mathcal{R}_i]} \tag{A.2}$$

where

$$\theta^{(m-1)\backslash *} = \{\phi_j, \psi_j, \alpha_{0,j}, \alpha_{1,j}, \beta_j; j = 1, \ldots, m, j \neq j^*\} \in \mathbb{R}^{2(m-1)} \times (\mathbb{R}^+)^{3(m-1)}$$
$$\theta^* = \{\phi^*_i, \psi^*_i, \alpha^*_{0,i}, \alpha^*_{1,i}, \beta^*_i; i \in \{j^*_{left}, j^*_{right}\}\} \in \mathbb{R}^4 \times (\mathbb{R}^+)^6$$

(II) Minimize the negative conditional pseudo log-likelihood implied by the refined partition $\mathcal{P}^{(m)}$ with respect to θ^*, when holding fixed the parameter vector $\hat{\theta}^{(m-1)\backslash *}$ implied by the previous partition $\mathcal{P}^{(m-1)}$:

$$\hat{\theta}^* = \arg\min_{\theta^*}(-\ell^{\mathcal{P}^{(m)}}(\hat{\theta}^{(m-1)\backslash *}, \theta^*; \mathbf{X}^n_2)) \tag{A.3}$$

In this step $-\ell^{\mathcal{P}^{(m)}}$ is implied by $g^{\mathcal{P}^{(m)}}(\cdot,\cdot,\cdot)$ and $f^{\mathcal{P}^{(m)}}(\cdot,\cdot,\cdot,\cdot)$ from (A.1) and (A.2), respectively, based on the refined partition $\mathcal{P}^{(m)}$. Starting values for θ^* in both new cells $\mathcal{R}_{j^*,left}$, $\mathcal{R}_{j^*,right}$ are given by the components of $\hat{\theta}^{(m-1)}$ associated with the partition cell[16] $\mathcal{R}_{j^*}$.

(III) For any possible partition of the form $\mathcal{P}^{(m)}$ compute the negative log-likelihood implied by the estimate (A.3) following steps (I) and (II) above. Select the optimal partition $\mathcal{P}^{(m)}_{opt}$ that attains the lowest negative log-likelihood.

3. Compute the pseudo maximum likelihood estimate $\hat{\theta}^{(m)}$ of the parameter $\theta^{(m)}$ associated with the partition $\mathcal{P}^{(m)}_{opt}$ by minimizing the negative pseudo log-likelihood (8) using functions $g^{\mathcal{P}^{(m)}_{opt}}$ and $f^{\mathcal{P}^{(m)}_{opt}}$ from (7). Starting values for this optimization are the parameter values $\hat{\theta}^{(m-1)\backslash *}$ associated with the previous optimal partition $\mathcal{P}^{(m-1)}_{opt}$ and the minimizer $\hat{\theta}^*$ in (A.3) which has been computed in step 2 above.

[16] The conditional pseudo log-likelihood (A.3) yields a substantial computational shortcut when compared with a full likelihood optimization. For instance, for every partition $\mathcal{P}^{(m)}$ in our estimations in Section 3 expression (A.3) involves only a 10-dimensional parameter θ^*. Since our algorithm searches over many candidate partitions $\mathcal{P}^{(m)}$ in every iteration step m, a relatively fast nonlinear minimization is important. The best partition $\mathcal{P}^{(m)}$ in every step m is determined by the maximal reduction of the negative conditional pseudo log-likelihood. We remark that the parameter estimates $\hat{\theta}^{(m)}$ in step 3 are computed from the full likelihood. This takes advantage of the fact that the starting values specified in step 3 are very reasonable in order to obtain a reliable and fast pseudo maximum likelihood estimate.

4. Repeat steps 2 and 3 until $m = M$. This yields a partition $\mathcal{P}_{opt}^{(M)}$ corresponding to a large binary tree equipped with parameter estimates $\hat{\theta}^{(M)}$.

REFERENCES

Audrino F, Bühlmann P. 2001. Tree-structured GARCH models. *Journal of the Royal Statistical Society, Series B* **63**(4): 727–744.

Audrino F, Bühlmann P. 2004. Synchronizing multivariate financial time series. *Journal of Risk* **6**(2): 81–106.

Bae K, Karolyi GA. 1994. Good news, bad news and international spillovers of stock return volatility between Japan and the US. *Pacific-Basin Finance Journal* **2**: 405–438.

Baillie RT, Bollerslev T, Mikkelsen HO. 1996. Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **74**: 3–30.

Beckaert G, Wu G. 2000. Asymmetric volatility and risk in equity markets. *The Review of Financial Studies* **13**: 1–42.

Becker KG, Finnerty JE, Friedman JE. 1995. Economic news and equity market linkages between the US and UK. *Journal of Banking & Finance* **19**: 1191–1210.

Black F. 1976. Studies in stock price volatility changes. *Proceedings of The American Statistical Association, Business and Economic Statistics Section*, 177–181.

Bollerslev T. 1986. Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics* **31**: 307–327.

Bollerslev T, Ghysels E. 1996. Periodic autoregressive conditional heteroskedasticity. *Journal of Business and Economic Statistics* **14**: 139–157.

Breiman L, Friedman JH, Olshen RA, Stone CJ. 1984. *Classification and Regression Trees*. Wadsworth: Belmont, CA.

Brockwell PJ, Davis RA. 1991. *Time Series: Theory and Methods*. Springer-Verlag: New York.

Chen CWS, Chiang TC, So MKP. 2003. Asymmetrical reaction to US stock-return news: evidence from major stock markets based on a double-threshold model. *Journal of Economics and Business* **55**: 487–502.

Chiang TC. 1998. Stock returns and conditional variance–covariance: evidence from Asian stock markets. In *Emerging Capital Markets: Financial and Investment Issues*, Choi JJ, Doukas JA (eds). Quorum Books: Westport, CN; 241–252.

Diebold FX, Inoue A. 2001. Long memory and regime switching. *Journal of Econometrics* **105**: 131–159.

Diebold FX, Mariano RS. 1995. Comparing predictive accuracy. *Journal of Business and Economic Statistics* **13**: 253–263.

Efron B, Tibshirani RJ. 1993. *An Introduction to the Bootstrap*. Chapman & Hall: London.

Engle RF. 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of U.K. inflation. *Econometrica* **50**: 987–1008.

Engle RF, Ito TR, Lin W. 1990. Meteor showers or heat waves? Heteroskedastic intra-daily volatility in the foreign exchange market. *Econometrica* **58**: 525–542.

Eun CS, Shin S. 1989. International transmission of stock market movements. *Journal of Financial and Quantitative Analysis* **24**: 241–256.

Fornari F, Mele A. 1997. Sign- and volatility-switching ARCH models: theory and applications to international stock markets. *Journal of Applied Econometrics* **12**: 49–65.

Glosten LR, Jagannathan R, Runkle D. 1993. On the relationship between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance* **48**: 1779–1801.

Gonzales-Rivera G. 1998. Smooth transition GARCH models. *Studies in Nonlinear Dynamics and Econometrics* **3**(2): 61–78.

Granger CWJ. 2001. Long memory processes—an economists viewpoint. Working paper, University of California, San Diego.

Granger CWJ, Hyung N. 2000. Occasional structural breaks and long memory. Working paper 99-14, University of California, San Diego.

Hamao Y, Masulis R, Ng V. 1990. Correlations in price changes and volatility across international stock markets. *The Review of Financial Studies* **3**: 281–307.

Haregud GE. 1997. A new non-linear GARCH model. PhD thesis, Stockholm School of Economics.
Hwang S, Satchell S. 1998. Implied volatility forecasting: a comparison of different procedures including fractionally integrated models with applications to UK equity options. In *Forecasting Volatility in the Financial Markets*, Knight J, Satchell S (eds). Butterworth: London.
Kim SW, Rogers JH. 1995. International stock price spillovers and market liberalization: evidence from Korea, Japan and the United States. *Journal of Empirical Finance* **2**: 117–133.
King MA, Wadhwani S. 1990. Transmission of volatility between stock markets. *The Review of Financial Studies* **3**: 5–33.
Koutmos G, Booth GG. 1995. Asymmetric volatility transmission in international stock markets. *Journal of International Money and Finance* **14**: 747–762.
Li CW, Li WK. 1996. On a double threshold autoregressive heteroscedastic time series model. *Journal of Applied Econometrics* **11**: 253–274.
Liu J, Li WK, Li CW. 1997. On a threshold autoregression with conditional heteroscedastic variances. *Journal of Statistical Planning and Inference* **62**(2): 279–300.
Masih R, Masih AMM. 2001. Long and short term dynamic causal transmission amongst international stock markets. *Journal of International Money and Finance* **20**: 563–587.
Medeiros MC, Veiga A. 2002. Are there multiple regimes in financial volatility? Working paper series, *Computing in Economics and Finance*, **311**.
Nelson DB. 1991. Conditional heteroscedasticity in asset returns: a new approach. *Econometrica* **59**: 347–370.
Nocedal J, Wright SJ. 1999. *Numerical Optimization*. Springer-Verlag: New York.
Pesaran MH, Timmermann A. 1992. A simple nonparametric test of predictive performance. *Journal of Business and Economic Statistics* **10**: 461–465.
Poon S, Granger C. 2003. Forecasting financial market volatility: a review. *Journal of Economic Literature* **41**(2): 478–539.
Rabemananjara R, Zakoian JM. 1993. Threshold ARCH models and asymmetries in volatility. *Journal of Applied Econometrics* **8**: 31–49.
Verhoeven P, McAleer M. 2001. Modeling conditional volatility using flexible smooth transition GARCH processes. In *Proceedings of the International Conference on Modelling and Forecasting Financial Volatility*, Vol. 3, Socioeconomic Systems, Franses PH, McAleer M, Chan F, Hoti S, Lim LK (eds). Australian National University: Canberra, Australia; 115–158.
Zakoian JM. 1994. Threshold heteroskedastic models. *Journal of Economic Dynamics and Control* **18**: 931–955.

