



# Multivariate fractionally integrated APARCH modeling of stock market volatility: A multi-country study<sup>☆</sup>

Christian Conrad<sup>a,\*</sup>, Menelaos Karanasos<sup>b,1</sup>, Ning Zeng<sup>c</sup>

<sup>a</sup> University of Heidelberg, Faculty of Economics and Social Studies, Bergheimer Str. 58, 69115 Heidelberg, Germany

<sup>b</sup> Economics and Finance, Brunel University, Uxbridge, West London, UB3 3PH, UK

<sup>c</sup> International Business School, Jinan University, Zhuhai 519070, China

## ARTICLE INFO

Available online 10 May 2010

JEL classification:

C13  
C22  
C32  
C52  
C53

Keywords:

Asymmetric power ARCH  
Fractional integration  
Stock returns  
Volatility forecast evaluation

## ABSTRACT

Tse (1998) proposes a model which combines the fractionally integrated GARCH formulation of Baillie, Bollerslev and Mikkelsen (1996) with the asymmetric power ARCH specification of Ding, Granger and Engle (1993). This paper analyzes the applicability of a multivariate constant conditional correlation version of the model to national stock market returns for eight countries. We find this multivariate specification to be generally applicable once power, leverage and long-memory effects are taken into consideration. In addition, we find that both the optimal fractional differencing parameter and power transformation are remarkably similar across countries. Out-of-sample evidence for the superior forecasting ability of the multivariate FIAPARCH framework is provided in terms of forecast error statistics and tests for equal forecast accuracy of the various models.

© 2010 Elsevier B.V. All rights reserved.

## 1. Introduction

A common finding in much of the empirical finance literature is that although the returns on speculative assets contain little serial correlation, the absolute returns and their power transformations are highly correlated. In particular, Ding et al. (1993) investigate the autocorrelation structure of  $|r_t|^\delta$ , where  $r_t$  is the daily S&P 500 stock market return, and  $\delta$  is a positive number. They find that  $|r_t|^\delta$  has significant positive autocorrelations for long lags. Motivated by this empirical result they propose a new general class of ARCH models, which they call the Asymmetric Power ARCH (APARCH). In addition, they show that this formulation comprises seven other specifications in the literature.<sup>2</sup> For an in depth discussion of the theoretical properties of the APARCH model see Karanasos and Kim (2006). McCurdy and Michaud (1996) and Tse (1998) extend the asymmetric power formulation of the variance to incorporate fractional integration, as defined by Baillie et al. (1996).<sup>3</sup> The new specification is termed fractionally integrated APARCH (FIAPARCH).

<sup>☆</sup> The final version of this paper was written while Ning Zeng was a PhD student at Brunel University. An earlier version of this paper by the first two authors circulated under the title "Fractionally Integrated APARCH Modeling of Stock Market Volatility: A multi-country study".

\* Corresponding author. Tel.: +49 6221 54 3173; fax: +49 6221 54 3578.

E-mail addresses: [christian.conrad@awi.uni-heidelberg.de](mailto:christian.conrad@awi.uni-heidelberg.de) (C. Conrad), [menelaos.karanasos@brunel.ac.uk](mailto:menelaos.karanasos@brunel.ac.uk) (M. Karanasos), [tingzeng@jnu.edu.cn](mailto:tingzeng@jnu.edu.cn) (N. Zeng).

<sup>1</sup> Tel.: +44 1895265284; fax: +44 1895269770.

<sup>2</sup> These models are: the ARCH (Engle, 1982), the GARCH (Bollerslev, 1986), the Taylor/Schwert GARCH in standard deviation (Taylor, 1986, and Schwert, 1990), the GJR GARCH (Glosten et al., 1993), the TARARCH (Zakoian, 1994), the NARCH (Higgins and Bera, 1992) and the log-ARCH (Geweke, 1986, and Pantula, 1986).

<sup>3</sup> The FIGARCH model of Baillie et al. (1996) is closely related to the long-memory GARCH process introduced in Karanasos et al. (2003) (see also Conrad and Karanasos, 2006, and Conrad, forthcoming, and the references therein).

The FIAPARCH model increases the flexibility of the conditional variance specification by allowing (a) an asymmetric response of volatility to positive and negative shocks, (b) the data to determine the power of returns for which the predictable structure in the volatility pattern is the strongest, and (c) long-range volatility dependence. These three features in the volatility processes of asset returns have major implications for many paradigms in modern financial economics. For example, the pricing of long-term options and optimal portfolio allocations must take into account all of these three properties.

At the same time, the FIAPARCH specification possesses the useful property that it nests the formulation without power effects and the stable one as special cases. This provides an encompassing framework for these two broad classes of specifications and facilitates comparison between them. The main contribution of this paper is to enhance our understanding of whether and to what extent this type of model improves upon its simpler counterparts.

Brooks et al. (2000) provide evidence for the applicability of the univariate APARCH model to national stock market returns for ten countries plus a world index. The results by Tse (1998) suggest that the FIAPARCH model is applicable to the yen–dollar exchange rate. More recently, Degiannakis (2004) and Níguez (2007) have applied univariate FIAPARCH specifications to stock return data. So far, multivariate versions of the framework have rarely been used in the literature. Only Dark (2004) applies a bivariate error correction FIAPARCH model to examine the relationship between stock and future markets, and Kim et al. (2005) use a bivariate FIAPARCH-in-mean process to model the volume–volatility relationship. Therefore, an interesting research issue is to explore how generally applicable this formulation is to a wide range of financial data and whether multivariate specifications can outperform their univariate counterparts. In this paper we address this issue by estimating both univariate and multivariate versions of this framework for eight series of national stock market index returns. These countries are Canada, France, Germany, Hong Kong, Japan, Singapore, the United Kingdom and the United States. As the general multivariate specification adopted in this paper nests the various univariate formulations, the relative ranking of each of these models can be considered using the Wald testing procedures and standard information criteria. Furthermore, the ability of the FIAPARCH formulation to forecast stock volatility out-of-sample is assessed by a variety of forecast error statistics. In order to verify whether the difference between the statistics from the various models is statistically significant we employ the tests of Diebold and Mariano (1995) and Harvey et al. (1997).

The remainder of this article is structured as follows. In Section 2 we detail the univariate and multivariate FIAPARCH models and discuss the various nested ARCH specifications. Section 3 discusses the data and presents the empirical results. In Section 4 we evaluate the different specifications in terms of their out-of-sample forecast ability. Finally, Section 5 concludes the analysis.

## 2. FIAPARCH model

### 2.1. Univariate process

One of the most common models in finance and economics to describe a time series  $r_t$  of stock returns is the AR(1) process

$$(1-\zeta L)r_t = c + \varepsilon_t, \quad t \in \mathbb{N}, \quad (2.1)$$

with

$$\varepsilon_t = e_t \sqrt{h_t},$$

where  $|c| \in [0, \infty)$ ,  $|\zeta| < 1$  and  $\{e_t\}$  are independently and identically distributed (*i.i.d.*) random variables with  $E(e_t) = E(e_t^2 - 1) = 0$ ,  $h_t$  is positive with probability one and is a measurable function of  $\Sigma_{t-1}$ , which in turn is the sigma-algebra generated by  $\{r_{t-1}, r_{t-2}, \dots\}$ . That is  $h_t$  denotes the conditional variance of the returns  $\{r_t\}$ , i.e.  $E[r_t | \Sigma_{t-1}] = c + \zeta r_{t-1}$  and  $\text{Var}[r_t | \Sigma_{t-1}] = h_t$ .

Tse (1998) examines the conditional heteroskedasticity of the yen–dollar exchange rate by employing the FIAPARCH (1,  $d$ , 1) model. Accordingly, we utilize the following process

$$(1-\beta L) \left( h_t^{\delta/2} - \omega \right) = \left[ (1-\beta L) - (1-\phi L)(1-L)^d \right] (1 + \gamma s_t) |\varepsilon_t|^\delta, \quad (2.2)$$

where  $\omega \in (0, \infty)$ ,  $|\beta| < 1$ ,  $|\phi| < 1$ ,  $0 \leq d \leq 1$ ,  $s_t = 1$  if  $\varepsilon_t < 0$  and 0 otherwise,  $\gamma$  is the leverage coefficient, and  $\delta$  is the parameter for the power term that takes (finite) positive values. A sufficient condition for the conditional variance  $h_t$  to be positive almost surely for all  $t$  is that  $\gamma > -1$  and the parameter combination  $(\phi, d, \beta)$  satisfies the inequality constraints provided in Conrad and Haug (2006) and Conrad (forthcoming).

When  $d=0$ , the process in Eq. (2.2) reduces to the APARCH(1,1) one, which nests two major classes of ARCH models. Specifically, a Taylor/Schwert type of formulation is specified when  $\delta=1$ , and a Bollerslev type is specified when  $\delta=2$ . There seems to be no obvious reason why one should assume that the conditional standard deviation is a linear function of lagged absolute returns or the conditional variance a linear function of lagged squared returns. As Brooks et al. (2000, p. 378) point out “the common use of a squared term in this role ( $\delta=2$ ) is most likely to be a reflection of the normality assumption traditionally invoked regarding financial data. However, if we accept that (high frequency) data are very likely to have a non-normal error

distribution, then the superiority of a squared term is lost and other power transformations may be more appropriate. Indeed, for non-normal data, by squaring the returns one effectively imposes a structure on the data which may potentially furnish sub-optimal modeling and forecasting performance relative to other power terms”.<sup>4</sup>

When  $\gamma = 0$  and  $\delta = 2$  the process in Eq. (2.2) reduces to the FIGARCH(1,  $d$ , 1) specification which includes Bollerslev’s (1986) model (when  $d = 0$ ) and the integrated specification (when  $d = 1$ ) as special cases. Baillie et al. (1996) point out that a striking empirical regularity that emerges from numerous studies of high-frequency, say daily, asset pricing data with ARCH-type models concerns the apparent widespread finding of integrated behavior. This property has been found in stock returns, exchange rates, commodity prices and interest rates (see Bollerslev et al., 1992). Yet unlike I(1) processes for the mean, there is less theoretical motivation for truly integrated behavior in the conditional variance. As noted by Baillie et al. (1996), for the variance being confined to only considering the extreme cases of stable and integrated specifications can be very misleading when long-memory (but eventually mean-reverting) processes are generating the observed data. They show that data generated from a process exhibiting long-memory volatility may be easily mistaken for integrated behavior.<sup>5</sup>

### 2.2. Multivariate formulation

Next, we introduce the multivariate version of the FIAPARCH specification. Let us define the  $N$ -dimensional column vector of the returns  $\mathbf{r}_t$  as  $\mathbf{r}_t = [r_{it}]_{i=1,\dots,N}$  and the corresponding residual vector  $\boldsymbol{\varepsilon}_t$  as  $\boldsymbol{\varepsilon}_t = [\varepsilon_{it}]_{i=1,\dots,N}$ . Next, the structure of the multivariate AR(1) mean equation is given by

$$\mathbf{Z}(L)\mathbf{r}_t = \mathbf{c} + \boldsymbol{\varepsilon}_t, \tag{2.3}$$

where  $\mathbf{c} = [c_i]_{i=1,\dots,N}$  with  $|c_i| \in [0, \infty)$  and  $\mathbf{Z}(L) = \text{diag}\{\zeta(L)\}$  is an  $N \times N$  diagonal matrix with  $\zeta(L) = [1 - \zeta_i L]_{i=1,\dots,N}$ ,  $|\zeta_i| < 1$ .<sup>6</sup> We assume that the noise vector  $\boldsymbol{\varepsilon}_t$  is characterized by the relation

$$\boldsymbol{\varepsilon}_t = \mathbf{e}_t \odot \mathbf{h}_t^{\wedge 1/2}, \tag{2.4}$$

where  $\odot$  and  $\wedge$  denote the Hadamard product and elementwise exponentiation respectively,  $\mathbf{h}_t = [h_{it}]_{i=1,\dots,N}$  is  $\Sigma_{t-1}$  measurable and the stochastic vector  $\mathbf{e}_t = [e_{it}]_{i=1,\dots,N}$  is independent and identically distributed with mean zero and positive definite covariance matrix  $\boldsymbol{\rho} = [\rho_{ij}]_{i,j=1,\dots,N}$  with  $\rho_{ij} = 1$  for  $i = j$ . From the above equation it follows that  $\mathbf{E}(\boldsymbol{\varepsilon}_t | \mathcal{F}_{t-1}) = \mathbf{0}$  and  $\mathbf{H}_t = \mathbf{E}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t' | \mathcal{F}_{t-1}) = \text{diag}\{\mathbf{h}_t^{\wedge 1/2}\} \boldsymbol{\rho} \text{diag}\{\mathbf{h}_t^{\wedge 1/2}\}$ . Hence,  $\mathbf{h}_t$  is the vector of conditional variances and  $\rho_{ij} = h_{ij,t} / \sqrt{h_{it} h_{jt}}$ ,  $i, j = 1, \dots, N$  are the constant conditional correlations (CCC).

Further, the multivariate FIAPARCH (M-FIAPARCH) process of order (1,  $d$ , 1) is defined by

$$\mathbf{B}(L) \left( \mathbf{h}_t^{\wedge \frac{\delta}{2}} - \boldsymbol{\omega} \right) = [\mathbf{B}(L) - \Delta(L) \Phi(L)] [\mathbf{I}_N + \boldsymbol{\Gamma}_t] |\boldsymbol{\varepsilon}_t|^{\wedge \delta}, \tag{2.5}$$

where  $|\boldsymbol{\varepsilon}_t|$  is the vector  $\boldsymbol{\varepsilon}_t$  with elements stripped of negative values. Moreover,  $\mathbf{B}(L) = \text{diag}\{\beta(L)\}$  with  $\beta(L) = [1 - \beta_i L]_{i=1,\dots,N}$ ,  $|\beta_i| < 1$ , and  $\Phi(L) = \text{diag}\{\phi(L)\}$  with  $\phi(L) = [1 - \phi_i L]_{i=1,\dots,N}$ ,  $|\phi_i| < 1$ . In addition,  $\boldsymbol{\omega} = [\omega_i]_{i=1,\dots,N}$  with  $\omega_i \in (0, \infty)$  and  $\Delta(L) = \text{diag}\{\mathbf{d}(L)\}$  with  $\mathbf{d}(L) = [(1-L)^{d_i}]_{i=1,\dots,N}$ ,  $0 \leq d_i \leq 1$ . Finally,  $\boldsymbol{\Gamma}_t = \text{diag}\{\gamma \odot \mathbf{s}_t\}$  with  $\gamma = [\gamma_i]_{i=1,\dots,N}$ , and  $\mathbf{s}_t = [s_{it}]_{i=1,\dots,N}$  where  $s_{it} = 1$  if  $\varepsilon_{it} < 0$  and 0 otherwise.

Note, that in our specification we do not allow for residual or volatility spillovers between the conditional variances. Because of this assumption, when  $\gamma_i > -1$ ,  $i = 1, \dots, N$ , the non-negativity conditions derived in Conrad (forthcoming) can be applied to each equation individually, which will ensure the positive definiteness of the conditional covariance matrix  $\mathbf{H}_t$  for all  $t$  almost surely. For results on the more general (short-memory) case which includes volatility spillovers, see Conrad and Karanasos (2010).

## 3. Empirical analysis

### 3.1. Data

Daily stock price index data for eight countries were sourced from the Datastream database for the period 1st January 1988 to 22nd April 2004, giving a total of 4,255 observations. We will use the period 1st January 1988 to 16th July 2003 for the estimation, while we produce 200 out-of-sample forecasts for the period 17th July 2003 to 22nd April 2004. The eight countries and their respective price indices are: UK: FTSE 100 (F), US: S&P 500 (SP), Germany: DAX 30 (D), France: CAC 40 (C), Japan: Nikkei 225 (N), Singapore: Straits Times (S), Hong Kong: Hang Seng (H) and Canada: TSE 300 (T). For each national index, the continuously

<sup>4</sup> For applications of the APARCH model in economics see Campos and Karanasos (2008) and Karanasos and Schurer (2008).

<sup>5</sup> An excellent survey of major econometric work on long-memory processes and their applications in economics and finance is given by Baillie (1996). For applications of the FIGARCH model to stock returns, interest rates, exchange rates and turnover volume see, (among others), Bollerslev and Mikkelsen (1996), Karanasos et al. (2006), Conrad and Lamla (2010) and Karanasos and Kartsaklas (2009), respectively.

<sup>6</sup> By  $\text{diag}(\mathbf{x})$  we denote a diagonal matrix with entries given by the elements of the vector  $\mathbf{x} = [x_i]_{i=1,\dots,N}$ .

compounded return was estimated as  $r_t = 100[\log(p_t) - \log(p_{t-1})]$  where  $p_t$  is the price on day  $t$ . A preliminary analysis of the squared return series based on 12th order Ljung–Box Q-statistics revealed high serial correlation in the second conditional moment of all indices. Furthermore, for all indices the Jarque–Bera statistic rejected the normality hypothesis at the 1% level. Finally, the estimated kurtosis coefficient is significantly above three for all indices but FTSE 100 and Nikkei 225. In order to accommodate the presence of such leptokurtosis, following Beine et al. (2002) and Níguez (2007) we assume student- $t$  distributed innovations  $e_t$  ( $\mathbf{e}_t$ ) and estimate the various specifications using the maximum likelihood estimation (MLE) method as implemented by Davidson (2008) in Time Series Modelling (TSM). For details on the likelihood function for the univariate and multivariate models see Davidson (2008). In addition, in order to check for the robustness of our results with respect to potential misspecification of the density, in Section 3.5 we apply a Gaussian density and quasi-maximum-likelihood estimation (QMLE). Because the student- $t$  distribution nests the normal distribution, this approach in turn facilitates the statistical comparison needed to discriminate between the two distributions.

### 3.2. Univariate models

We proceed with the estimation of the AR(1)-FIAPARCH(1,  $d$ , 1) model in Eqs. (2.1) and (2.2) in order to take into account the serial correlation and the GARCH effects observed in our time series data, and to capture the possible long-memory in volatility. The only exceptions are the Canadian and Singaporean indices, where an AR(1)-FIAPARCH(0,  $d$ , 1) model is used. For these two indices the AR(1)-FIAPARCH(1,  $d$ , 1) estimates for  $\beta$  were insignificant and the Akaike and Schwarz information criteria (AIC and SIC) came out in favor of the (0,  $d$ , 1) specification. In addition, for the Hang Seng index the criteria favor the (1,  $d$ , 0) formulation.

Table 1 reports the estimation results. For reasons of brevity, we do not present the estimates of the constants in the mean and the variance, which were significant in all cases but one. In all countries the AR coefficient ( $\zeta$ ) is highly significant. As mentioned above, the estimate for the  $\phi(\beta)$  parameter is insignificant only in one (two) out of the eight cases. In three countries the estimates of the leverage term ( $\gamma$ ) are statistically significant, confirming the hypothesis that there is negative correlation between returns and volatility. For the other countries we reestimated the models without an asymmetry term. For all indices the estimates of the power term ( $\delta$ ) and the fractional differencing parameter ( $d$ ) are highly significant. Interestingly, the highest power terms are obtained for the two American indices, while the European ones are characterized by the highest degree of persistence. In all cases, the estimated degrees of freedom parameter ( $\nu$ ) is highly significant and leads to an estimate of the kurtosis which is different from three.<sup>7</sup>

In all cases, the ARCH parameters satisfy the set of conditions which guarantee the non-negativity of the conditional variance (see Conrad (forthcoming)). According to the values of the Ljung–Box tests for serial correlation in the standardized and squared standardized residuals there is no statistically significant evidence of misspecification in almost all cases.

#### 3.2.1. Tests of fractional differencing and power term parameters

A large number of studies have documented the persistence of volatility in stock returns; see, e.g., Ding et al. (1993) and Ding and Granger (1996). Using daily data many of these studies have concluded that the volatility process is well approximated by an IGARCH specification. However, from the FI(A)PARCH estimates reported in Table 1, it appears that the long-run dynamics are better modeled by the fractional differencing parameter. To test for the persistence of the conditional heteroskedasticity models, we examine the Wald statistics for the linear constraints  $d = 0$  (stable APARCH) and  $d = 1$  (IAPARCH).<sup>8</sup> For reasons of brevity we omit the table with the test results, which are available from the authors upon request. In summary, the Wald tests clearly reject both the stable and integrated null hypotheses against the FIAPARCH one for all indices. Thus, purely from the perspective of searching for a model that best describes the volatility in the stock return series, the fractionally integrated one appears to be the most satisfactory representation.<sup>9</sup>

This result is an important finding because the time series behavior of volatility affects asset prices through the risk premium. Christensen and Nielsen (2007) establish theoretically and empirically the consequences of long-memory in volatility for asset prices. Using a model for expected returns to discount streams of expected future cash flows, they calculate asset prices. Within this context the risk-return trade-off and the serial correlation in volatility are the two most important determinants of asset values. Christensen and Nielsen (2007) derive the way in which these two ingredients jointly determine the level of stock prices. They also investigate the quantitative economic consequences of long-memory in volatility on asset price elasticities and show that the elasticity is smaller in magnitude than earlier estimates, and much more stable under variations in the long-memory parameter than in the short-memory case. Thus, they point out that the high elasticities reported earlier should be interpreted with considerable caution. They also highlight the fact that the way in which volatility enters in the asset evaluation model is crucial and should be considered carefully. This is due to the fact that the memory properties of the volatility process carry over to the stock return process through the risk premium link (see also Christensen et al., 2010 and Conrad et al., 2010).

<sup>7</sup> The kurtosis of a student- $t$  distributed random variable with  $\nu$  degrees of freedom is  $3(\nu - 2)/(\nu - 4)$ .

<sup>8</sup> Restricting  $d$  to be zero in Eq. (2.2) leads to an APARCH(1,1) model with parameters  $\beta$  and  $\phi - \beta$ . Similarly, restricting  $d$  to be one that leads to an IAPARCH(1,2) model with parameters  $\beta$ ,  $1 + \phi - \beta$  and  $-\phi$ .

<sup>9</sup> It is worth mentioning the empirical results in Granger and Hyung (2004). They suggest that there is a possibility that at least part of the long-memory may be caused by the presence of neglected breaks in the series. We look forward to clarifying this in future work.

**Table 1**  
Univariate AR-FI(A)PARCH models (MLE).

	SP	T	C	D	F	H	N	S
$\zeta$	−0.05 <sup>a</sup> (−3.28)	0.17 (10.94)	0.04 (2.34)	0.03 <sup>a</sup> (2.31)	0.04 (2.38)	0.06 (3.71)	−0.02 (−1.63)	0.15 (9.20)
$\beta$	0.54 (5.81)	–	0.66 (6.94)	0.56 (5.65)	0.59 (5.32)	0.08 (2.01)	0.51 (4.83)	–
$\phi$	0.27 (4.11)	−0.11 (−2.95)	0.20 (3.84)	0.21 (4.82)	0.19 (3.77)	–	0.14 (2.03)	−0.07 (−2.23)
$\gamma$	–	–	–	0.46 (3.73)	–	0.69 (3.65)	–	0.76 (3.90)
$\delta$	2.35 (23.50)	2.42 (17.28)	1.77 (12.64)	1.24 (11.46)	1.86 (14.31)	1.28 (12.80)	2.07 (18.81)	1.40 (12.73)
$d$	0.30 (6.00)	0.19 (6.33)	0.52 (4.33)	0.40 (4.34)	0.46 (4.60)	0.18 (4.50)	0.42 (6.00)	0.21 (5.25)
$\nu$	5.60 (10.77)	5.38 (10.76)	8.53 (6.56)	6.83 (6.90)	10.70 (6.04)	4.56 (11.12)	5.80 (10.54)	4.86 (11.04)
$Q_{12}$	18.45 [0.10]	9.52 [0.66]	10.00 [0.61]	13.18 [0.36]	12.86 [0.38]	22.85 [0.03]	10.59 [0.56]	18.50 [0.10]
$Q_{12}^2$	5.12 [0.95]	19.47 [0.08]	11.74 [0.47]	8.13 [0.77]	18.00 [0.12]	33.24 [0.00]	20.90 [0.05]	2.20 [1.00]

Notes: For each of the eight indices, Table 1 reports MLE results for the AR(1)-FI(A)PARCH model. The numbers in parentheses are *t*-statistics.  $Q_{12}$  and  $Q_{12}^2$  are the 12th order Ljung–Box tests for serial correlation in the standardized and squared standardized residuals respectively. The numbers in brackets are *p*-values.

<sup>a</sup> For the S&P 500 and Dax 30 indices we estimate AR(3) and AR(4) models respectively.

In order to check for the robustness of the Wald testing results discussed above, we apply the Akaike, Schwarz, Hannan–Quinn or Shibata information criteria (AIC, SIC, HQIC, and SHIC respectively) to rank the various ARCH type models. Specifically, according to the AIC, HQIC and SHIC, the optimal specification (i.e., FIAPARCH, APARCH or IAPARCH) for all indices was the FIAPARCH one (values not reported).<sup>10</sup> The SIC results are largely in line with the AIC, HQIC or SHIC results.

Next, recall that the two common values of the power term imposed throughout much of the GARCH literature are the values of two (Bollerslev's model) and unity (the Taylor/Schwert specification). The invalid imposition of a particular value for the power term may lead to sub-optimal modeling and forecasting performance (Brooks et al., 2000). Accordingly, we test whether the estimated power terms are significantly different from unity or two using Wald tests (results not reported). We find that all eight estimated power coefficients are significantly different from unity. Further, with the exception of the CAC 40, FTSE 100 and Nikkei 225 indices, each of the power terms is significantly different from two. Hence, on the basis of these results, in the majority of cases support is found for the (asymmetric) power fractionally integrated model, which allows an optimal power transformation term to be estimated. The evidence obtained from the Wald tests is reinforced by the model ranking provided by the four model selection criteria (values not reported). This is a noteworthy result since He and Teräsvirta (1999) emphasized that if the standard Bollerslev type of model is augmented by the 'heteroskedasticity' parameter, the estimates of the ARCH and GARCH coefficients almost certainly change. More importantly, Karanasos and Schurer (2008) show that in the univariate GARCH-in-mean level formulation the significance of the in-mean effect is sensitive to the choice of the power term.

### 3.3. Multivariate models

The analysis above suggests that the FIAPARCH formulation describes the conditional variances of the eight stock indices well. However, financial volatilities move together over time across assets and markets. Recognizing this commonality through a multivariate modeling framework can lead to obvious gains in efficiency compared to working with separate univariate specifications (Bauwens and Laurent, 2005). Therefore, multivariate GARCH models are essential for enhancing our understanding of the relationships between the (co)volatilities of economic and financial time series. For recent surveys on multivariate specifications and their practical importance in various areas such as asset pricing, portfolio selection and risk management see e.g., Bauwens et al. (2006) and Silvennoinen and Teräsvirta (2007). Thus in this section, within the framework of the multivariate CCC model, we will analyze the dynamic adjustments of the variances for the various indices. Overall we estimate seven bivariate specifications; three for the European countries: CAC 40–DAX 30 (C–D), CAC 40–FTSE 100 (C–F) and DAX 30–FTSE 100 (D–F); three for the Asian countries: Hang Seng–Nikkei 225 (H–N), Hang Seng–Straits Times (H–S) and Nikkei 225–Straits Times (N–S); one for the S&P 500 and TSE 300 indices (SP–T). Moreover, we estimate two trivariate models: one for the three European countries (C–D–F) and one for the three Asian countries (H–N–S).

#### 3.3.1. Bivariate processes

The best fitting bivariate specification is chosen according to likelihood ratio results and the minimum value of the information criteria (not reported). In the majority of the models the AR coefficients are significant at the 5% level or better. In almost all cases a

<sup>10</sup> Caporin (2003) performs a Monte Carlo simulation study and verifies that information criteria clearly detect the presence of long-memory in volatility.

(1,  $d$ , 1) order is chosen for the FIAPARCH formulation. Only for the H–S and N–S models do we choose (0,  $d$ , 1) order for the Straits Times index, and (1,  $d$ , 0) order for the Hang Seng index. Note that this is in line with our findings for the univariate models, where the  $\beta$  parameter was insignificant for Straits Times, while the  $\phi$  parameter was insignificant for Hang Seng. In six out of the fourteen models the leverage term ( $\gamma$ ) is significant. As in the univariate case, it is significant in both indices for the H–S case and in the DAX 30 index for the D–F case. In addition, in the bivariate case it is also significant in the TSE 300 index for the SP–T model and in the Nikkei 225 for the N–S one. In almost all cases the power term ( $\delta$ ) and the fractional differencing parameter ( $d$ ) are highly significant. In the D–F, H–S and N–S models the two countries generated very similar power terms: (1.28, 1.36), (1.42, 1.47) and (1.70, 1.62) respectively. In four out of the seven bivariate formulations the two countries generated very similar fractional parameters. These are the SP–T, the C–F, the H–N and the H–S models. The corresponding pairs of values are: (0.22, 0.21), (0.24, 0.29), (0.36, 0.35) and (0.16, 0.13). Interestingly, in the majority of the cases the estimated power and fractional differencing parameters of the bivariate models take lower values than those of the corresponding univariate models. In all cases the estimated CCC ( $\rho$ ) is highly significant. Interestingly, it is rather high among the American and European indices, and rather low among the Asian indices. Finally, the degrees of freedom parameters ( $\nu$ ) are highly significant and the ARCH parameters satisfy the Conrad (forthcoming) conditions. In the majority of the cases the hypothesis of uncorrelated standardized and squared standardized residuals is well supported (see the last two rows of Table 2).

Next, for each pair of indices from Table 2, we examine the Wald statistics for the linear constraints  $d_i = d_j = 0$  (stable APARCH) and  $d_i = d_j = 1$  (IAPARCH). For all indices the Wald tests clearly reject both the stable and integrated null hypotheses against the FIAPARCH one (results not reported). We also test whether the estimated power terms are significantly different from unity or two using Wald tests. The eight estimated power coefficients are significantly different from either unity or two (again the results are not reported).

### 3.3.2. Trivariate specifications

Table 3 reports the parameters of interest for the two trivariate FI(A)PARCH(1, $d$ ,1) models.<sup>11</sup> In two out of the three Asian countries the leverage term ( $\gamma$ ) is weakly significant. In all cases the power term ( $\delta$ ) and the fractional differencing parameter ( $d$ ) are highly significant. Similarly, in all cases the estimated CCC ( $\rho$ ) and degrees of freedom parameters ( $\nu$ ) are highly significant and the ARCH parameters satisfy the conditions provided in Conrad (forthcoming). In particular, the estimates of  $\rho$  confirm the results from the bivariate models, i.e. the conditional correlation between the European indices is considerably stronger than between the Asian indices.

### 3.4. On the similarity of the fractional/power parameters

Next, we test for the apparent similarity of the optimal fractional differencing and power term parameters for each of the eight country indices using pairwise Wald tests:

$$W_d = \frac{(d_1 - d_2)^2}{\text{Var}(d_1) + \text{Var}(d_2) - 2\text{Cov}(d_1, d_2)} \quad \text{and} \quad W_\delta = \frac{(\delta_1 - \delta_2)^2}{\text{Var}(\delta_1) + \text{Var}(\delta_2) - 2\text{Cov}(\delta_1, \delta_2)},$$

where  $d_i$  ( $\delta_i$ ),  $i = 1, 2$ , is the fractional differencing (power term) parameter from the bivariate FIAPARCH model estimated for the national stock market index for country  $i$ ,  $\text{Var}(d_i)$ , and  $\text{Var}(\delta_i)$  are the corresponding variances, and  $\text{Cov}(d_1, d_2)$ , and  $\text{Cov}(\delta_1, \delta_2)$  are the corresponding covariances. The above Wald statistics test whether the fractional differencing (power term) parameters of the two countries are equal  $d_1 = d_2$  ( $\delta_1 = \delta_2$ ), and are distributed as  $\chi^2_{(1)}$ .

The following table presents the results of this pairwise testing procedure for the various bivariate models.<sup>12</sup> Several findings emerge from this table. The estimated long-memory parameters for the various (a)symmetric specifications are in the range  $0.20$  ( $0.13 \leq d \leq 0.48$ ) ( $0.36$ ) while the estimated power terms are in the range  $1.19$  ( $1.18 \leq \delta \leq 2.00$ ) ( $1.86$ ). In all cases for the American and Asian indices (and in the majority of the cases for the European countries) the values of the two coefficients ( $d_i$ ,  $\delta_i$ ) for the asymmetric models are lower than the corresponding values for the symmetric formulations. The values of the Wald tests in the table support the null hypothesis that the two estimated fractional parameters and the two power term coefficients are not significantly different from one another.

All specifications generated very similar long-memory coefficients between countries. For example, in the asymmetric SP–T and H–N models, which generated very similar fractional parameters (0.22, 0.23 and 0.36, 0.35 respectively), the two coefficients were, as expected, not significantly different ( $W = 0.04, 0.02$  respectively). The null hypothesis of equal long-memory coefficients is rejected at the 5% level only for the symmetric C–D and the asymmetric D–F models. Both include the DAX 30 index with a relatively high persistence parameter. As regards the power term, the two models for CAC 40 and DAX 30 indices are those with the highest differences:  $1.59 - 1.18 = 0.41$  and  $1.55 - 1.19 = 0.36$  respectively. For these two cases the values of the Wald tests ( $W = 6.57, 3.85$  respectively) are significant at the 5% level. For all other models but one, the equality of the power terms cannot be rejected. For example, in models which generated very similar power terms, such as the symmetric D–F one (1.35, 1.40) or the asymmetric H–S (1.42, 1.47), the two coefficients were, as expected, not significantly different ( $W = 0.10$  in both cases). Finally, it

<sup>11</sup> The complete results are of course available upon request.

<sup>12</sup> For reasons of comparability, in all the various bivariate models for both indices we estimated AR(1)-FI(A)PARCH(1,  $d$ , 1) processes. That is, the parameter values for  $d$  and  $\delta$  presented in Table 4 are not necessarily the same as the ones in Table 2.

**Table 2**  
Bivariate AR-FI(A)PARCH models (MLE).

	SP–T		C–D		C–F		D–F		H–N		H–S		N–S	
	SP	T	C	D	C	F	D	F	H	N	H	S	N	S
$\zeta_i$	−0.05 <sup>a</sup> (−4.51)	0.17 (13.86)	−0.03 (−2.63)	0.02 <sup>a</sup> (1.53)	0.05 (3.88)	0.04 (2.70)	0.01 <sup>a</sup> (0.37)	−0.03 (−2.14)	0.05 (3.44)	−0.02 (−1.07)	0.03 (1.79)	0.14 (9.16)	−0.03 (−2.08)	0.15 (9.33)
$\beta_i$	0.46 (4.78)	0.33 (2.27)	0.50 (3.94)	0.62 (9.00)	0.35 (1.55)	0.45 (1.48)	0.55 (5.51)	0.42 (1.93)	0.57 (3.70)	0.46 (3.82)	0.08 (2.85)	–	0.43 (3.41)	–
$\phi_i$	0.26 (3.73)	0.18 (1.52)	0.26 (4.30)	0.24 (5.60)	0.16 (1.24)	0.20 (1.48)	0.20 (4.96)	0.17 (1.74)	0.33 (3.94)	0.15 (2.11)	–	−0.02 (0.87)	0.14 (1.76)	−0.07 (1.78)
$\gamma_i$	–	0.34 (2.46)	–	–	–	–	0.14 (1.68)	–	–	0.10 (1.74)	0.11 (1.73)	0.47 (3.16)	0.11 (2.02)	–
$\delta_i$	1.85 (8.81)	1.59 (8.37)	1.55 (9.12)	1.23 (9.84)	1.76 (7.65)	1.55 (5.54)	1.28 (11.64)	1.36 (8.00)	1.49 (17.71)	1.69 (13.75)	1.42 (12.07)	1.47 (12.01)	1.70 (13.75)	1.62 (15.71)
$d_i$	0.22 (5.50)	0.21 (5.25)	0.30 (3.00)	0.44 (6.28)	0.24 (2.18)	0.29 (1.61)	0.40 (4.44)	0.28 (2.15)	0.36 (3.18)	0.35 (5.04)	0.16 (7.58)	0.13 (5.79)	0.33 (5.15)	0.23 (6.64)
$\rho$		0.65 (21.33)		0.65 (20.54)		0.67 (20.90)		0.54 (19.48)		0.33 (11.03)		0.43 (17.02)		0.26 (12.32)
$v$		13.69 (9.85)		16.69 (6.76)		18.96 (6.94)		18.13 (6.06)		12.62 (11.03)		11.31 (11.44)		12.42 (10.47)
$Q_{12}$	18.08 [0.11]	10.74 [0.55]	34.92 [0.00]	10.17 [0.60]	10.33 [0.59]	15.80 [0.20]	12.48 [0.41]	36.27 [0.00]	22.30 [0.03]	9.78 [0.63]	39.18 [0.00]	20.69 [0.06]	12.93 [0.37]	16.36 [0.17]
$Q_{12}^2$	2.77 [0.99]	2.81 [0.99]	20.28 [0.06]	5.17 [0.95]	24.51 [0.02]	40.18 [0.00]	3.31 [0.99]	17.44 [0.13]	35.79 [0.00]	54.52 [0.00]	108.29 [0.00]	5.03 [0.96]	58.83 [0.00]	1.50 [1.00]

Notes: For each of the seven pairs of indices, Table 2 reports MLE results for the bivariate AR-FI(A)PARCH model. SP–T denotes the bivariate process for the S&P 500 and TSE 300 indices. C–D, C–F and D–F indicate the three bivariate models for the European indices. H–N, H–S and N–S stand for the three bivariate specifications for the Asian indices. The numbers in parentheses are  $t$ -statistics.  $Q_{12}$  and  $Q_{12}^2$  are the 12th order Ljung–Box tests for serial correlation in the standardized and squared standardized residuals respectively. The numbers in brackets are  $p$ -values.

<sup>a</sup> For the S&P 500 and DAX 30 indices we estimate AR models of orders 3 and 4 respectively.

is noteworthy that in the majority of cases the values of the coefficients  $d_i$  and  $\delta_i$  for the univariate (a)symmetric formulations are higher than the corresponding values for the (a)symmetric bivariate and trivariate models (not reported).

### 3.5. Robustness

In order to check whether our findings are robust to the choice of the error distribution, we reestimate all specifications using a Gaussian density and QMLE. The results are presented in Tables 7–9 in the Appendix. The univariate, bivariate and trivariate QMLE results are very similar to the ones based on student- $t$  distributed innovations. Most importantly, the estimated fractional differencing parameters and power terms are close to the corresponding values estimated under the student- $t$  distribution and MLE. In particular, the hypotheses that  $d = 0$ ,  $d = 1$  as well as the one that  $\delta = 1$  are clearly rejected for all specifications and all indices with only one exception. In contrast to the results under the student- $t$  distribution, in the case of QMLE the hypothesis that  $\delta = 2$  cannot be rejected in most of the cases. Also the evidence for asymmetries becomes slightly weaker. However, the

**Table 3**  
Trivariate AR-FI(A)PARCH(1,  $d$ , 1) models (MLE).

	C–D–F			H–N–S			
	C	D	F	H	N	S	
$\beta_i$	0.19 (1.40)	0.43 (4.61)	0.22 (1.15)	0.39 (1.92)	0.38 (2.50)	0.78 (18.85)	
$\phi_i$	0.11 (0.90)	0.22 (3.35)	0.09 (0.61)	0.28 (1.56)	0.15 (1.53)	0.81 (22.08)	
$\gamma_i$	–	–	–	0.02 (1.46)	0.07 (1.60)	–	
$\delta_i$	1.83 (10.95)	1.25 (9.52)	1.56 (7.12)	1.47 (13.36)	1.53 (10.20)	1.88 (11.75)	
$d_i$	0.11 (4.16)	0.25 (5.43)	0.15 (3.27)	0.18 (4.50)	0.26 (0.07)	0.08 (4.39)	
	C–D		D–F	C–F	H–N	N–S	H–S
$\rho$	0.66 (21.07)		0.56 (19.86)	0.68 (21.70)	0.32 (14.84)	0.25 (12.19)	0.43 (16.92)
$v$			9.60 (17.36)			8.42 (20.54)	

Notes: Table 3 reports MLE results for the two trivariate FI(A)PARCH(1,  $d$ , 1) models. C–D–F and H–N–S denote the models for the European and Asian countries respectively. The numbers in parentheses are  $t$ -statistics.

**Table 4**

Tests for similarity of fractional and power terms (Bivariate Models).

	Symmetric Models							Asymmetric models						
	SP-T	C-D	C-F	D-F	H-N	H-S	N-S	SP-T	C-D	C-F	D-F	H-N	H-S	N-S
$d$														
$d_1$	0.24	0.30	0.24	0.48	0.38	0.26	0.34	0.22	0.19	0.26	0.36	0.36	0.16	0.32
$d_2$	0.27	0.45	0.29	0.41	0.36	0.20	0.24	0.23	0.29	0.33	0.23	0.35	0.13	0.22
$W$	0.25	4.16	0.26	0.75	0.04	0.85	1.46	0.04	1.24	0.05	6.00	0.02	1.61	1.62
$\delta$														
$\delta_1$	2.00	1.55	1.76	1.35	1.50	1.49	1.80	1.86	1.59	1.74	1.27	1.49	1.42	1.66
$\delta_2$	1.68	1.19	1.55	1.40	1.79	1.68	1.68	1.51	1.18	1.51	1.39	1.70	1.47	1.58
$W$	2.38	3.85	1.08	0.10	4.43	1.59	0.60	2.59	6.57	1.55	0.24	1.66	0.10	0.19

Notes: SP-T denotes the bivariate model for the S&P 500 and TSE 300 indices respectively. C-D, C-F and D-F indicate the three bivariate models for the European indices. H-N, H-S and N-S stand for the three bivariate models for the Asian indices. The  $W$  rows report the corresponding Wald statistics. The 5% and 1% critical values are 3.84 and 6.63 respectively.

information criteria again clearly favor the most general FIAPARCH specification (not reported). Hence, the results from the QMLE reinforce our previous findings and highlight the importance of allowing for flexible fractional differencing parameters and power terms. Nevertheless, the information criteria and likelihood ratio tests almost unambiguously favor the models estimated under the student- $t$  distribution and MLE over the ones under normality and QMLE (results are not reported).

## 4. Forecasting methodology

### 4.1. Evaluation criteria

In Section 3 we showed that on the basis of several model selection techniques the superior *in-sample* fitting specification was the FIAPARCH one. However, for practical forecasting purposes, the predictive ability of these models needs to be examined *out-of-sample*. The aim of this section is to examine the relative ability of the various univariate and multivariate long-memory and power formulations to forecast daily stock return volatility.

Several empirical studies examine the forecast performance of various GARCH models (see, e.g., Hansen and Lunde, 2005). The survey by Poon and Granger (2003) provides, among other things, an interesting and extensive synopsis of them. Since the publication of Ding et al. (1993) there has been a great deal of research investigating if the fractional integrated models could help to make better volatility forecasts.<sup>13</sup> Hyung et al. (2008) compare the out-of-sample forecasting performance of various short and long-memory volatility models. They find that for forecast horizons of 10 days and beyond, the FIGARCH specification is the dominant one. The long-memory characteristic has important implications for volatility forecasting and option pricing. Option pricing in a stochastic volatility setting requires a risk premium for the unhedgeable volatility risk. The fractionally integrated series lead to volatility forecasts larger than those from short-memory models, which immediately translate into higher option prices. This could be an explanation for the better pricing performance of the FIGARCH model in this case (Hyung et al., 2008).

Our full sample consists of 4255 trading days and each model is estimated over the first 4055 observations of the full sample, i.e. over the period 1st January 1988 to 16th July 2003. As a result the out-of-sample period is from 17th July 2003 to 22nd April 2004, providing 200 daily observations. The parameter estimates obtained with the data from the in-sample period are inserted in the relevant forecasting formulas and volatility forecasts  $\hat{h}_{t+1}$  calculated given the information available at time  $t = T (= 4055), \dots, T + 199 (= 4254)$ , i.e. 200 one-step ahead forecasts are calculated.

In order to evaluate the forecast performance of the different model specifications we need (a) to obtain a valid proxy for the true but unobservable underlying volatility and (b) to specify certain loss functions. A natural candidate for the proxy are the squared returns, which are an unbiased estimator for the unobserved conditional variance. However, compared to realized volatility the squared returns are a noisy proxy and, as shown in Patton (forthcoming), distortions in the rankings of competing forecasts can arise when using noisy proxies. Whether such distortions arise depends on the choice of the loss function. Patton (forthcoming) provides necessary and sufficient conditions on the functional form of the loss function to ensure that the ranking is the same whether it is based on the true conditional variance or some conditionally unbiased volatility proxy. Two loss functions which satisfy these conditions are the mean square error (MSE) statistic and the QLIKE statistic. Consequently, we will employ the MSE and the QLIKE statistic, where the latter corresponds to the loss implied by a Gaussian likelihood. Finally, in addition to those robust loss functions we make use of the adjusted mean absolute percentage error (AMAPE) statistic (see Table 5 below). In contrast to the simple mean absolute percentage error the AMAPE corrects for the problem of asymmetry between the actual and forecast values.

<sup>13</sup> For the literature on the forecasting performance of univariate fractionally integrated and power ARCH models see, among others, Degiannakis (2004), Hansen and Lunde (2006) and Níguez (2007).



**Table 5**  
Forecast evaluation criteria.

MSE	$k^{-1} \sum_{t=T+1}^{T+k} (\hat{h}_t - r_t^2)^2$
QLIKE	$k^{-1} \sum_{t=T+1}^{T+k} [\ln(\hat{h}_t) + r_t^2 / \hat{h}_t]$
AMAPE	$k^{-1} \sum_{t=T+1}^{T+k}  (\hat{h}_t - r_t^2) / (\hat{h}_t + r_t^2) $

Notes:  $k$  is the number of one-step ahead forecasts,  $T$  is the sample size,  $\hat{h}_t$  is the forecasted variance and  $r_t^2$  are the squared returns.

**Table 6**  
Best versus worst ranked models.

	MSE	QLIKE	AMAPE
S&P 500	B-FIAP vs. U-FIAP [0.00]	B-IAP vs. U-FIAP [0.03]	B-AP vs. U-FIAP [0.02]
TSE 300	B-FIAP vs. U-IAP [0.14]	U-FIP vs. U-IAP [0.00]	B-AP vs. U-IAP [0.00]
CAC 40	T-P vs. B <sub>F</sub> -FIA( $\delta=2$ ) [0.00]	T-IP vs. B <sub>F</sub> -FIA( $\delta=2$ ) [0.15]	T-IP vs. B <sub>F</sub> -FIA( $\delta=2$ ) [0.00]
DAX 30	B <sub>F</sub> -AP vs. U-FIAP [0.00]	U-FIA( $\delta=1$ ) vs. B <sub>C</sub> -FIA( $\delta=2$ ) [0.08]	B <sub>F</sub> -AP vs. B <sub>F</sub> -FIA( $\delta=2$ ) [0.17]
FTSE 100	T-P vs. B <sub>C</sub> -FIA( $\delta=2$ ) [0.00]	T-P vs. B <sub>C</sub> -FIA( $\delta=2$ ) [0.01]	B <sub>D</sub> -AP vs. B <sub>C</sub> -FIA( $\delta=2$ ) [0.00]
Hang Seng	B <sub>S</sub> -FIA vs. U-AP [0.00]	B <sub>N</sub> -AP vs. T-FIAP [0.02]	T-FIA( $\delta=2$ ) vs. U-FIA( $\delta=2$ ) [0.26]
Nikkei 225	B <sub>S</sub> -FIA( $\delta=1$ ) vs. U-FIAP [0.12]	U-FI( $\delta=1$ ) vs. T-AP [0.03]	T-FIA( $\delta=2$ ) vs. U-AP [0.67]
Straits Times	B <sub>H</sub> -FIAP vs. B <sub>N</sub> -IAP [0.00]	B <sub>H</sub> -FIA( $\delta=2$ ) vs. U-AP [0.01]	T-FIAP vs. U-AP [0.00]

Notes: U, B and T stand for univariate, bivariate and trivariate specifications respectively. (F)I, A and P indicate (fractionally) integrated, asymmetric and power models respectively. The subscripts refer to the jointly estimated index of the bivariate model, e.g., the subscript  $F$  indicates that the bivariate model is estimated with the FTSE 100 index. The numbers in brackets are the  $p$ -values from the Diebold and Mariano (1995) test.

For each index we calculated the three forecast error statistics for the specifications APARCH, IAPARCH, FIAPARCH( $\delta=1$ ), FIAPARCH( $\delta=2$ ) and FIAPARCH in the univariate, bivariate and (where possible) trivariate versions. Hence, overall, fifteen values of each forecast error statistic are available for each index. Instead of presenting all the figures, we decided to present in Table 6 only the best and the worst specifications for each index as identified by the forecast error statistics. In addition, we test whether the values of the forecast error statistics from the best and the worst model are statistically significant using the Diebold and Mariano (1995) test. Table 6 contains the corresponding  $p$ -values (see the next section).

An examination of Table 6 reveals that either a multivariate or a fractionally integrated (FI) or a power (P) or an asymmetric (A) process is clearly superior. That is, there is strong evidence that the restrictive univariate (U), stable, symmetric Bollerslev's type of process is inferior to one of the more flexible specifications. The results can be summarized as follows. Only in three cases is the best ranked model, as assessed by the forecasting criteria, the univariate one. Both MSE and AMAPE loss functions uniformly favor either bivariate or trivariate specifications (see the second and fourth columns of Table 6). For the two American indices in five out of the six cases a bivariate model is selected as being best (see the first two rows of Table 6). Similarly, the results for the European countries show the close connection between the three volatilities. In five cases a trivariate specification is the best performing model and in three cases a bivariate one. For the Asian indices in only one case do the statistics rank the univariate formulation first (see the last three rows of Table 6). Overall, the multivariate formulation has the best statistics for twenty one out of the twenty four cases.

Moreover, in the Asian countries the (fractionally) integrated model is favored in all but one case. Similarly, for the S&P 500 and the TSE 300 indices the statistics indicate the superiority of the fractionally integrated specification. The power formulation is the dominant one in the European and American countries. In particular, for the European indices the restriction that  $\delta=2$  characterizes, with one exception, the worst performing specification. In summary, the best formulations as ranked by the forecast error statistics are multivariate models. For the American and Asian indices the long-memory property appears to be important for the forecast performance, while for the European and American indices power specifications are dominant.

#### 4.2. Tests of equal forecast accuracy

In some cases the error statistics do not allow for a clear distinction between the ranked models, which is evidenced by the marginal difference in relative accuracy which separates the models (results are not reported). Thus next we move to the pairwise comparison of the best and the worst specifications.

We utilize the tests proposed by Diebold and Mariano (1995) and Harvey et al. (1997). Before moving to the two tests some notation is needed. First, we denote the one-step ahead loss functions for the best and worst models as  $L_{bt}^{(i)}(r_t^2, \hat{h}_{bt})$  and  $L_{wt}^{(i)}(r_t^2, \hat{h}_{wt})$  ( $t = T + 1, \dots, T + k$ ), where  $i \in \{\text{MSE, QLIKE, AMAPE}\}$ , respectively. Forecasts of the squared returns are generated using the fixed forecasting scheme (described in West and McCracken, 1998, p. 819). Next, let  $\Delta_t = L_{bt}^{(i)} - L_{wt}^{(i)}$  and  $\bar{\Delta}$  denote its sample mean, i.e.  $\bar{\Delta} = k^{-1} \sum_{t=T+1}^{T+k} \Delta_t$ . The test proposed by Diebold and Mariano (1995) is formed as

$$S = \left[ \widehat{\text{Var}}(\bar{\Delta}) \right]^{-1/2} \bar{\Delta}, \quad \text{with} \quad \widehat{\text{Var}}(\bar{\Delta}) = \frac{2\pi \hat{f}_{\Delta}(0)}{k},$$

where  $\hat{f}_{\Delta}(0)$  is a consistent estimate of the spectral density function of  $\Delta$  at frequency zero. Under the null hypothesis  $S$  has an asymptotic standard normal distribution.<sup>14</sup>

As seen in Table 6 the evidence obtained from the loss functions is reinforced by the Diebold–Mariano test. Clearly the test discriminates between the best and the worst model. That is, in the majority of the cases (eighteen out of twenty four) the test indicates the superiority of the best formulation over the worst one. In particular, for the USA and Canada, in four out of the five cases the worst model (univariate) is rejected in favor of the best (multivariate) one. For the Asian indices, the Diebold–Mariano test indicates the superiority of the best (fractionally integrated) specification over the worst (stable) one in four out of the five cases. Further, for the European countries, in five out of the seven cases the power (best) formulation outperforms the Bollerslev (worst) one.<sup>15</sup>

## 5. Conclusion

In this paper strong evidence has been put forward suggesting that the conditional volatility of eight national stock indices is best modeled as a FIAPARCH process. On the basis of Wald tests and information criteria the fractionally integrated model provides a statistically significant improvement over its integrated counterpart. One can also reject the more restrictive stable process, and consequently all the existing specifications (see Ding et al. 1993) nested by it in favor of the fractionally integrated parameterization. Hence, our analysis has shown that the FIAPARCH formulation is preferred to both the stable and the integrated ones.

Additionally, according to our analysis, all eight countries show strong evidence of power effects when long-memory persistence in the conditional volatility has been taken into account, as both the Bollerslev and Taylor/Schwert specifications were rejected in favor of the power formulation. Further, comparing the pairwise testing results of the log-likelihood procedures to the relative model rankings provided by the four alternative criteria we observed that the findings were generally robust. That is, where the log-likelihood results provided unanimous support for the FIAPARCH specification over either the Bollerslev or Taylor/Schwert (asymmetric) FIGARCH formulations, the model selection criteria were compatible without exception. Thus, the inclusion of a power term and a fractional unit root in the conditional variance equation appear to augment the model in a worthwhile fashion.

We would also like to emphasize that the above results were robust to the dimension of the process. That is, the evidence on the superiority of the FIAPARCH specification obtained from the univariate models was reinforced by the multivariate processes. Moreover, the apparent similarity of the fractional differencing and power terms suggests that the M-FIAPARCH model has a quite general empirical validity across many different markets. Finally, our in-sample results are reinforced by a comparison of the out-of-sample forecast performance of the various models, which shows that the M-FIAPARCH specification is generally preferred to its nested competitors.

Finally, we would like to point to two potentially interesting extensions of the multivariate models studied in this article, namely specifications with slowly time-varying intercepts as suggested in Baillie and Morana (2009) or with volatility spillovers as considered in Conrad and Karanasos (2010).

## Acknowledgments

We are very grateful to Richard Baillie, the editor, and two anonymous referees for their detailed comments which led to a substantial improvement on the previous version of this article. We would also like to thank Markus Haas, Marika Karanassou and the participants of the annual meeting of the Verein für Socialpolitik (Magdeburg, 2009) for their helpful suggestions.

<sup>14</sup> Harvey et al. (1997) propose a small sample correction for the Diebold and Mariano (1995) statistic. Their modified test statistic is student- $t$  distributed with  $k - 1$  degrees of freedom. The results from this statistic are qualitatively similar to the original Diebold and Mariano (1995) statistic and, hence, are not reported.

<sup>15</sup> We also utilize two encompassing tests proposed by Ericsson (1992) and Harvey et al. (1998). We do not report the results for reasons of brevity. We find that for the FTSE 100 index, in the univariate and bivariate F-C models, the FIAPARCH formulation outperforms the restricted Taylor/Schwert and Bollerslev specifications, and the stable/integrated ones as well.

## Appendix. QMLE results

Table 7

Univariate AR-FI(A)PARCH models (QMLE).

	SP	T	C	D	F	H	N	S
$\zeta$	−0.04 <sup>a</sup> (−2.64)	0.18 (10.44)	0.04 (2.69)	0.02 <sup>a</sup> (1.04)	0.05 (2.89)	0.10 (4.90)	−0.01 (−0.11)	0.16 (8.97)
$\beta$	0.49 (1.81)	–	0.48 (3.03)	0.59 (4.96)	0.47 (1.84)	0.24 (1.44)	0.38 (2.27)	0.12 (1.09)
$\phi$	0.23 (1.44)	−0.18 (−2.32)	0.18 (2.57)	0.13 (2.44)	0.18 (1.45)	–	0.10 (0.86)	–
$\gamma$	–	–	–	–	–	–	–	–
$\delta$	2.18 (4.95)	2.20 (13.75)	1.89 (10.50)	1.59 (7.23)	1.89 (9.95)	1.80 (12.00)	2.07 (15.92)	1.85 (11.56)
$d$	0.30 (2.00)	0.24 (3.00)	0.36 (2.57)	0.53 (3.53)	0.37 (2.31)	0.38 (2.71)	0.37 (4.62)	0.30 (2.73)
$Q_{12}$	18.11 [0.11]	8.02 [0.78]	10.24 [0.59]	11.55 [0.48]	12.26 [0.42]	15.57 [0.21]	8.10 [0.78]	12.70 [0.39]
$Q_{12}^2$	4.01 [0.98]	14.28 [0.28]	10.30 [0.59]	2.68 [1.00]	13.53 [0.33]	36.16 [0.00]	7.76 [0.80]	0.64 [1.00]

Notes: For each of the eight indices, Table 7 reports QMLE results for the AR(1)-FI(A)PARCH model. The numbers in parentheses are *t*-statistics.  $Q_{12}$  and  $Q_{12}^2$  are the 12th order Ljung–Box tests for serial correlation in the standardized and squared standardized residuals respectively. The numbers in brackets are *p*-values.

<sup>a</sup> For the S&P 500 and Dax 30 indices we estimate AR(3) and AR(4) models respectively.

Table 8

Bivariate AR-FI(A)PARCH models (QMLE).

	SP–T		C–D		C–F		D–F		H–N		H–S		N–S	
	SP	T	C	D	C	F	D	F	H	N	H	S	N	S
$\zeta_i$	−0.06 <sup>a</sup> (−4.74)	0.17 (13.33)	−0.03 (−2.67)	0.01 <sup>a</sup> (0.84)	0.06 (4.10)	0.05 (3.23)	−0.003 <sup>a</sup> (−0.17)	−0.02 (−1.63)	0.02 (1.11)	−0.01 (−0.63)	0.03 (1.71)	0.13 (7.77)	−0.02 (−1.28)	0.15 (8.33)
$\beta_i$	0.37 (2.70)	0.26 (1.30)	0.48 (3.65)	0.57 (5.45)	0.12 (2.11)	0.12 (1.81)	0.50 (3.13)	0.38 (0.93)	0.79 (8.67)	0.24 (3.53)	–	–	0.23 (3.48)	−0.84 (−8.74)
$\phi_i$	0.17 (1.61)	0.08 (0.47)	0.22 (2.87)	0.19 (3.70)	–	–	0.18 (3.44)	0.16 (0.69)	0.25 (1.64)	–	−0.13 (−2.32)	–	–	−0.87 (−10.26)
$\gamma_i$	–	0.33 (1.52)	–	–	–	–	0.16 (1.38)	–	–	0.12 (1.58)	–	–	0.16 (1.86)	–
$\delta_i$	1.99 (6.22)	1.77 (6.55)	1.61 (10.06)	1.35 (7.94)	2.04 (9.71)	1.91 (7.96)	1.37 (8.06)	1.47 (5.25)	1.69 (11.27)	2.09 (16.08)	1.84 (12.27)	2.03 (11.94)	1.97 (14.07)	1.84 (10.22)
$d_i$	0.23 (4.60)	0.25 (4.17)	0.34 (3.40)	0.47 (3.61)	0.17 (3.40)	0.19 (3.17)	0.40 (2.5)	0.29 (1.53)	0.72 (3.27)	0.31 (5.17)	0.25 (6.25)	0.16 (5.33)	0.30 (5.00)	0.22 (4.40)
$\rho$	0.67 (15.42)		0.65 (13.60)		0.67 (19.43)		0.54 (17.51)		0.35 (12.99)		0.49 (11.28)		0.30 (11.04)	
$Q_{12}$	18.39 [0.10]	10.28 [0.59]	36.62 [0.00]	11.38 [0.50]	11.34 [0.50]	15.39 [0.22]	14.37 [0.28]	31.78 [0.00]	46.36 [0.00]	10.26 [0.59]	34.93 [0.00]	18.09 [0.11]	10.90 [0.54]	13.73 [0.32]
$Q_{12}^2$	2.60 [1.00]	2.71 [1.00]	14.62 [0.26]	2.94 [1.00]	36.32 [0.00]	50.13 [0.00]	2.39 [1.00]	15.58 [0.21]	16.17 [0.18]	6.89 [0.86]	59.92 [0.00]	3.08 [0.99]	8.41 [0.75]	1.68 [1.00]

Notes: For each of the seven pairs of indices, Table 8 reports QMLE results for the bivariate AR-FI(A)PARCH model. SP–T denotes the bivariate process for the S&P 500 and TSE 300 indices. C–D, C–F and D–F indicate the three bivariate models for the European indices. H–N, H–S and N–S stands for the three bivariate specifications for the Asian indices. The numbers in parentheses are *t*-statistics.  $Q_{12}$  and  $Q_{12}^2$  are the 12th order Ljung–Box tests for serial correlation in the standardized and squared standardized residuals respectively. The numbers in brackets are *p*-values.

<sup>a</sup> For the S&P 500 and DAX 30 indices we estimate AR models of order 3 and 4 respectively.

Table 9

Trivariate AR-FI(A)PARCH(1, *d*, 1) models (QMLE).

	C–D–F			H–N–S		
	C	D	F	H	N	S
$\beta_i$	0.34 (1.99)	0.54 (4.60)	0.55 (3.18)	0.74 (4.93)	0.33 (1.74)	0.82 (15.59)
$\phi_i$	0.16 (1.53)	0.20 (3.74)	0.23 (3.06)	0.30 (0.72)	0.08 (0.54)	0.89 (21.78)
$\gamma_i$	–	–	–	0.02 (1.46)	0.10 (1.60)	–
$\delta_i$	1.70 (8.10)	1.36 (7.56)	1.39 (8.69)	1.57 (14.27)	1.98 (12.38)	2.26 (7.06)
$d_i$	0.24 (2.66)	0.43 (3.07)	0.40 (2.86)	0.60 (1.86)	0.32 (4.57)	0.06 (3.00)

(continued on next page)

Table 9 (continued)

	C–D–F			H–N–S		
	C	D	F	H	N	S
	C–D	D–F	C–F	H–N	N–S	H–S
$\rho$	0.66 (13.84)	0.55 (17.61)	0.67 (19.85)	0.35 (13.00)	0.30 (10.96)	0.50 (11.22)

Notes: Table 9 reports QMLE results for the two trivariate FI(A)PARCH(1,  $d$ , 1) models. C–D–F and H–N–S denote the models for the European and Asian countries respectively. The numbers in parentheses are  $t$ -statistics.

## References

- Baillie, R., 1996. Long memory processes and fractional integration in econometrics. *J. Econometrics* 73, 5–59.
- Baillie, R., Morana, C., 2009. Modeling long memory and structural breaks in conditional variances: an adaptive FIGARCH approach. *J. Econ. Dyn. Control* 33, 1577–1592.
- Baillie, R.T., Bollerslev, T., Mikkelsen, H.O., 1996. Fractionally integrated generalized autoregressive conditional heteroskedasticity. *J. Econometrics* 74, 3–30.
- Bauwens, L., Laurent, S., 2005. A new class of multivariate skew densities, with application to generalized autoregressive conditional heteroscedasticity models. *J. Bus. Econ. Stat.* 23, 346–354.
- Bauwens, L., Laurent, S., Rombouts, J.V.K., 2006. Multivariate GARCH models. *J. Appl. Econometrics* 21, 79–109.
- Beine, M., Bénassy-Quéré, A., Lecourt, C., 2002. Central bank intervention and foreign exchange rates: new evidence from FIGARCH estimations. *J. Int. Money Finance* 21, 115–144.
- Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. *J. Econometrics* 31, 307–327.
- Bollerslev, T., Mikkelsen, H.O., 1996. Modeling and pricing long memory in stock market volatility. *J. Econometrics* 73, 151–184.
- Bollerslev, T., Chou, R.Y., Kroner, K.F., 1992. ARCH modeling in finance: a review of the theory and empirical evidence. *J. Econometrics* 52, 5–59.
- Brooks, R.D., Faff, R.W., McKenzie, M.D., 2000. A multi-country study of power ARCH models and national stock market returns. *J. Int. Money Finance* 19, 377–397.
- Campos, N., Karanasos, M., 2008. Growth, volatility and political instability: non-linear time-series evidence for Argentina, 1890–2000. *Econ. Lett.* 100, 135–137.
- Caporin, B., 2003. Identification of long memory in GARCH models. *Stat. Meth. Appl.* 12, 133–151.
- Christensen, B.J., Nielsen, M.Ø., 2007. The effect of long memory in volatility on stock market fluctuations. *Rev. Econ. Stat.* 89, 684–700.
- Christensen, B.J., Nielsen, M.Ø., Zhu, J., 2010. Long memory in stock market volatility and the volatility-in-mean effect: the FIGARCH-M model. *J. Empirical Finance* 17, 460–470.
- Conrad, C., forthcoming. Non-negativity conditions for the hyperbolic GARCH model. *J. Econometrics*. doi:10.1016/j.jeconom.2010.03.045.
- Conrad, C., Haag, B., 2006. Inequality constraints in the fractionally integrated GARCH model. *J. Financ. Econometrics* 3, 413–449.
- Conrad, C., Karanasos, M., 2006. The impulse response function of the long memory GARCH model. *Econ. Lett.* 90, 34–41.
- Conrad, C., Karanasos, M., 2010. Negative volatility spillovers in the unrestricted ECCG-GARCH model. *Econometric Theory* 26, 838–862.
- Conrad, C., Lamla, M.J., 2010. The high-frequency response of the EUR-USD exchange rate to ECB communication. *Journal of Money, Credit and Banking*, forthcoming.
- Conrad, C., Elmer, S., Karanasos, M., 2010. On the transmission of memory: inflation persistence and the Great Moderation. Unpublished manuscript, University of Heidelberg.
- Dark, J., 2004. Bivariate error correction FIGARCH and FIAPARCH models on the Australian All Ordinaries Index and its SPI futures. Monash University, Unpublished paper.
- Davidson, J., 2008. Time series modelling 4.26. <http://www.timeseriesmodelling.com>.
- Degiannakis, S., 2004. Volatility forecasting: evidence from a fractional integrated asymmetric power ARCH skewed-t model. *Appl. Financ. Econ.* 14, 1333–1342.
- Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. *J. Bus. Econ. Stat.* 13, 253–263.
- Ding, Z., Granger, C.W.J., 1996. Modeling volatility persistence of speculative returns: a new approach. *J. Econometrics* 73, 185–215.
- Ding, Z., Granger, C.W.J., Engle, R.F., 1993. A long memory property of stock market returns and a new model. *J. Empirical Finance* 1, 83–106.
- Engle, R.F., 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica* 50, 987–1008.
- Ericsson, N.R., 1992. Parameter constancy, mean square forecast errors, and measuring forecast performance: an exposition, extension and illustration. *J. Policy Model.* 14, 465–495.
- Geweke, J., 1986. Modeling the persistence of conditional variances: a comment. *Econometric Rev.* 5, 57–61.
- Glosten, L.R., Jagannathan, Runkle, D., 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. *J. Finance* 48, 1779–1801.
- Granger, C., Hyung, N., 2004. Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns. *J. Empirical Finance* 11, 399–421.
- Hansen, P.R., Lunde, A., 2005. A forecast comparison of volatility models: does anything beat a GARCH(1, 1). *J. Appl. Econometrics* 20, 873–889.
- Hansen, P.R., Lunde, A., 2006. Consistent ranking of volatility models. *J. Econometrics* 131, 97–121.
- Harvey, D., Leybourne, S.J., Newbold, P., 1997. Testing the equality of prediction mean squared errors. *Int. J. Forecasting* 13, 281–291.
- Harvey, D., Leybourne, S.J., Newbold, P., 1998. Tests for forecast encompassing. *J. Bus. Econ. Stat.* 16, 254–259.
- He, C., Teräsvirta, T., 1999. Statistical properties of the asymmetric power ARCH model. In: Engle, R.F., White, H. (Eds.), *Cointegration, Causality and Forecasting*. Festschrift in honour of Clive W. J. Granger. Oxford University Press, Oxford, pp. 462–474.
- Higgins, M., Bera, A., 1992. A class of nonlinear ARCH models. *Int. Econ. Rev.* 33, 137–158.
- Hyung, N., Poon, S.-H., Granger, C., 2008. A source of long memory in volatility. In: Rapach, D.E., Wohar, M.E. (Eds.), *Forecasting in the Presence of Structural Breaks and Model Uncertainty*. Emerald, pp. 329–380.
- Karanasos, M., Kartsaklas, A., 2009. Dual long-memory, structural breaks and the link between turnover and the range-based volatility. *J. Empirical Finance* 16, 838–851.
- Karanasos, M., Kim, J., 2006. A re-examination of the asymmetric power ARCH model. *J. Empirical Finance* 13, 113–128.
- Karanasos, M., Schurer, S., 2008. Is the relationship between inflation and its uncertainty linear? *Ger. Econ. Rev.* 9, 265–286.
- Karanasos, M., Psaradakis, Z., Sola, M., 2003. On the autocorrelation properties of long-memory GARCH processes. *J. Time Ser. Anal.* 25, 265–281.
- Karanasos, M., Sekioua, S., Zeng, N., 2006. On the order of integration of monthly US ex-ante and ex-post real interest rates: new evidence from over a century of data. *Econ. Lett.* 90, 163–169.
- Kim, J., Kartsaklas, A., Karanasos, M., 2005. The volume–volatility relationship and the opening of the Korean stock market to foreign investors after the financial turmoil in 1997. *Asia-Pac. Financ. Mark.* 12, 245–271.
- McCurdy, T. H., Michaud, P. K., 1996. Capturing long memory in the volatility of equity returns: a fractionally integrated asymmetric power ARCH model. University of Toronto, unpublished manuscript.

- Ñíguez, T.-M., 2007. Volatility and VaR forecasting in the Madrid stock exchange. *Span. Econ. Rev.* 10, 169–196.
- Pantula, S., 1986. Modeling the persistence in conditional variances: a comment. *Econometric Rev.* 5, 71–74.
- Patton, A. J., forthcoming. Volatility forecast comparison using imperfect volatility proxies. *J. Econometrics*. doi:10.1016/j.jeconom.2010.03.034.
- Poon, S.H., Granger, C.W.J., 2003. Forecasting volatility in financial markets: a review. *J. Econ. Lit.* XLI, 478–539.
- Schwert, W., 1990. Stock volatility and the crash of '87. *Rev. Financ. Stud.* 3, 77–102.
- Silvennoinen, A., Teräsvirta, T., 2007. Multivariate GARCH models. In: Andersen, T.G., Davis, R.A., Kreiß, J.P., Mikosch, T. (Eds.), *Handbook of Financial Time Series*. Springer, New York.
- Taylor, S., 1986. *Modeling Financial Time Series*. Wiley, New York.
- Tse, Y.K., 1998. The conditional heteroscedasticity of the yen-dollar exchange rate. *J. Appl. Econometrics* 13, 49–55.
- West, D., McCracken, M.W., 1998. Regression-based tests of predictive ability. *Int. Econ. Rev.* 39, 817–840.
- Zakoian, J.M., 1994. Threshold heteroskedastic model. *J. Econ. Dyn. Control* 18, 931–995.