Lux, Thomas; Kaizoji, Taisei

Working Paper

Forecasting volatility and volume in the Tokyo stock market: The advantage of long memory models

Economics working paper / Christian-Albrechts-Universität Kiel, Department of Economics, No. 2004,05

Provided in cooperation with:
Institut für Weltwirtschaft (IfW)

Suggested citation: Lux, Thomas; Kaizoji, Taisei (2004) : Forecasting volatility and volume in the Tokyo stock market: The advantage of long memory models, Economics working paper / Christian-Albrechts-Universität Kiel, Department of Economics, No. 2004,05, http://hdl.handle.net/10419/3244

Forecasting Volatility and Volume in the Tokyo Stock Market: The Advantage of Long Memory Models

by Thomas Lux and Taisei Kaizoji

C|A|U
Christian-Albrechts-Universität Kiel
Department of Economics

Economics Working Paper
No 2004-05

Forecasting Volatility and Volume in the Tokyo Stock Market: The Advantage of Long Memory Models*

Thomas Lux [1] and Taisei Kaizoji [2]

[1] Department of Economics, University of Kiel,
Olshausenstr. 40, 24118 Kiel, Germany
lux@bwl.uni-kiel.de

[2] Institute of Social Sciences, International Christian University
Mitaka City, Tokyo 181 – 8585, Japan
kaizoji@icu.ac.jp

Abstract: We investigate the predictability of both volatility and volume for a large sample of Japanese stocks. The particular emphasis of this paper is on assessing the performance of long memory time series models in comparison to their short-memory counterparts. Since long memory models should have a particular advantage over long forecasting horizons, we consider predictions of up to 100 days ahead. In most respects, the long memory models (ARFIMA, FIGARCH and the recently introduced multifractal models) dominate over GARCH and ARMA models. However, while FIGARCH and ARFIMA also have a number of cases with dramatic failures of their forecasts, the multifractal model does not suffer from this shortcoming and its performance practically always improves upon the naïve forecast provided by historical volatility. As a somewhat surprising result, we also find that, for FIGARCH and ARFIMA models, pooled estimates (i.e. averages of parameter estimates from a sample of time series) give much better results than individually estimated models.

Keywords: forecasting, long memory models, volume, volatility

JEL-classification: C22, C53, G12

May 2004

Corresponding author:

Thomas Lux, e-mail: lux@bwl.uni-kiel.de, Tel. +49 431 880 3661, Fax: +49 431 880 4383

* We are grateful to seminar participants at the University of Kiel for helpful discussions and to Laurie Davis and Philipp Sibbertsen for stimulating exchanges on long-memory volatility models.

Introduction

It is well known that both volatility and trading volume are characterized by a much higher degree of predictability than the returns of financial assets. In the huge literature on forecasting volatility, the vast majority of papers use variants of the GARCH family of stochastic processes, which provide an easy and convenient way to capture the basic auto-regressive structure of conditional variances (see Granger and Poon, 2003, for a recent survey of the voluminous literature on volatility forecasting). However, results are not unanimously in favour of the potential of GARCH models to improve upon the forecasting performance of simpler models like the historical mean volatility or moving average or smoothed representations of it. Dimson and Marsh (1990) and Figlewski (1997), among others, find that simpler models can indeed outperform GARCH or related approaches at least when applied to low-frequency (weekly, monthly) data. On the other hand, dozens of papers investigate whether improvements over GARCH as a benchmark are possible using non-linear models or artificial intelligence techniques (e.g. West and Cho, 1995; Brailsford and Faff, 1996; Donaldson and Kamstra, 1997; Klaasen, 2002; Neely and Weller, 2002). However, given the ample evidence for long-term dependence in volatility (i.e. hyperbolic decay of its autocorrelation function rather than the exponential decay characteristic of short-memory models), it also appears worthwhile to explore the potential value added by models sharing this feature. Long memory in volatility has been first pointed out as a stylised fact by Ding, Engle and Granger (1992). Prevalence of this feature in financial data has meanwhile been confirmed in many subsequent studies and counts now as one of the truly universal properties of asset markets (cf. Lobato and Savin, 1998).

Long memory generalizations of standard short-memory time series models are available in the ARFIMA (Granger and Joyeux, 1980) and FIGARCH models (Baillie et al., 1996). When browsing the literature on volatility forecasting, it comes as a certain surprise that these candidate models have received relatively scant attention so far. Basically, only two papers with a direct comparison between GARCH and FIGARCH forecasts appear to be available at present, Vilasuso (2002) and Zumbach (2004) both considering volatility forecasts in foreign exchange markets. Vilasuso reports relatively large reductions of both mean squared error and mean absolute errors over forecasting horizons of 1 to 10 days with FIGARCH compared to GARCH. Zumbach's result using intra-daily data are more sobering in that he finds improvements in daily forecasts to be only of the order of one to two percent of MSE. Given that there is essentially only one study supportive of superior predictability of long-memory models, a more systematic analysis of this issue seems to be worthwhile.

The exclusive focus on exchange markets also raises the question how long-memory models would perform in other types of markets, e.g. in national stock markets. We attempt to shed light on both issues with our investigation of Japanese stocks. Our data base consists of daily prices and volume for more than 1,000 stocks traded in the first section of the Tokyo stock exchange (trading in the first section requires that certain criteria are met on outstanding shares, trading volume and dividend payments). Data are available at daily frequency over the twenty-seven years period 01/01/1975 to 12/31/2001. A rather typical example of the evolution of stock price and daily trading volume is shown in *Fig*. 1 for the Nippon Suisan Company. What stands out here and in most other time series is the enormous increase of stock prices during the Japanese bubble in the second half of the eighties and the decline thereafter. Another common feature of many stocks is the increase of trading volume during the bubble reaching levels that are often by far the highest over the whole sample. After the collapse of the bubble, prices gradually fell to their levels in the early eighties while volume also went down to roughly its level of the pre-bubble period. For some stocks in our data base, trading stopped before 2001 because of bankruptcy of the company. To our knowledge, analyses of the volatility dynamics of Japanese stocks have so far been confined to short memory GARCH type models (e.g., Tse, 1991; Fong, 1997). An exception is Ray *et al.* (1997) who estimate ARFIMA models for the Tokyo Stock Price Index (TOPIX) and 15 individual stocks. However, they are interested in the predictability of raw returns (they find predictable components in the range between 5 and 15% of monthly variations), but do not explore the issue of predicting volatility or volume as we will do in the following.

Since analysis of all stocks appeared to be too time-consuming, we selected two subsets of one hundred entries, respectively. The first of these subsets consisted of 100 randomly chosen stocks, the second of those with the largest average volume. For all these stocks we estimated four time series models for their volatility dynamics: GARCH, FIGARCH, ARFIMA and the 'causal multi-fractal model' recently introduced by Calvet and Fisher (2001), another model that at least allows to 'mimic' long-term dependence (see below for details). We included ARFIMA models to see whether a difference exists between the performance of the original ARFIMA structure applied to volatility and its embedding into a GARCH framework (i.e., the FIGARCH model). The multi-fractal model (a variant of the one proposed originally by Mandelbrot, Calvet and Fisher, 1997) provides an alternative formalization of long-term dependence in volatility and has already been found to outperform GARCH and FIGARCH in some time series (Calvet and Fisher, 2003; Lux, 2003). In contrast to the additive structure of the GARCH dynasty, the multi-fractal model conceives volatility as a hierarchical, multiplicative process with heterogeneous components, but, in fact, achieves this in a rather parsimonious way using (in the version applied here) only two parameters. Our overall finding is that improvements over GARCH can be achieved by alternative models which is in contrast to frequent findings of the opposite in the literature (which, however, mostly does not include long-memory models as alternatives).

Since volume is known to share the long memory property of volatility (Bollerslev and Jubinski, 1999; Ray and Tsay, 2000) and to be strongly contemporaneously correlated with volatility, it seems to be worthwhile to also investigate its predictability along similar lines. Since GARCH type models are not applicable to volume, we use only the ARFIMA and multi-fractal models and compare their performance to ARMA models as a short-memory benchmark. Again, dominance of the alternative models is confirmed. As it turns also out, the predictable component in volume appears to be much higher on average than that in volatility judged by the improvements in mean-squared errors and mean absolute errors against naïve forecasts.

In a final exercise, we use pooled parameters estimates (averages of the parameter estimates obtained by any particular model over the whole sample of 100 stocks) for forecasting of future volatility. Counterintuitively, discarding information about individual time series in this way leads to vast improvements of average forecast quality for volatility and, albeit to a lesser extent, for volume as well.

Our study proceeds along the following lines: sec. 2 deals with volatility forecasts. In sec. 2.1 we introduce the models and estimation methods, while results are presented in secs. 2.2 and 2.3. Similarly, models and results for volume are found in secs. 3.1 and 3.2./3.3. Sec. 4, then, considers pooled estimates. Our conclusions are to be found in sec. 5.

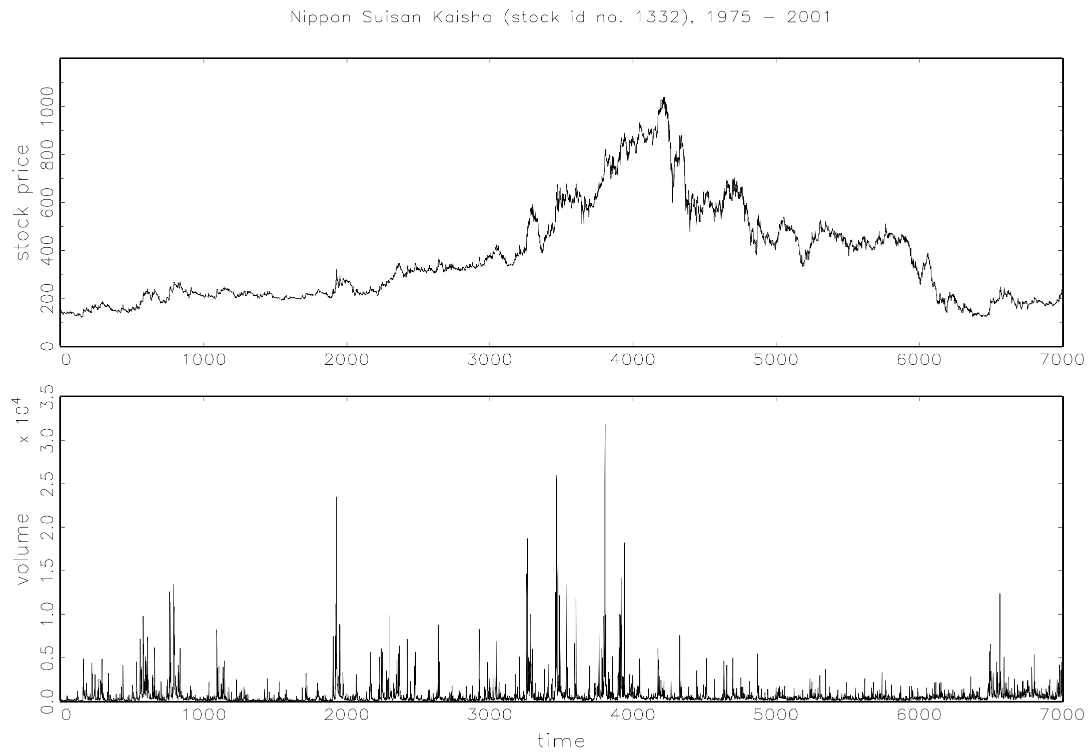

Fig. 1: Stock price and volume of Nippon Suisan Company (stock identification number 1332). Nippon Suisan is a company processing marine food. It has been established in 1911 and, as of September 2003, it had 1,534 employees and 36,336 shareholders. As with most other entries in our database, the Japanese bubble in the second half of the eighties has also affected the price evolution of this stock. As one typically observes, volume also reaches its maximum levels during the bubble phase.

1. Forecasting Volatility of Japanese Stocks

2.1. Models

In our analysis of forecastability of volatility, the standard benchmark is the GARCH (1,1) model, which we expand – like all other models – by allowing for a constant and first-order autoregressive component in raw returns (x_t):

$$(1)\quad x_t = \mu + \rho \cdot x_{t-1} + \sqrt{h_t} \cdot \varepsilon_t \qquad \text{with } \varepsilon_t \sim N(0, 1)$$

with volatility dynamics being governed by:

$$(2)\quad h_t = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}, \qquad \omega > 0,\ \alpha_1,\ ß_1 \geq 0.$$

The fractionally integrated extension of the GARCH model (FIGARCH) expands the variance equation by considering fractional differences. Like with the baseline GARCH model, we restrict attention to one lag in both the autoregressive and moving average terms, i.e., FIGARCH(1,d,1), which can be written as:

$$(3)\quad h_t = \omega + \beta_1 h_{t-1} + (1 - \beta_1 L - (1 - \varphi_1 L)(1 - L)^d)\varepsilon_t^2 .$$

As is well known, the Binomial expansion of the fractional difference operator introduces an infinite number of past lags with hyperbolically decaying coefficients. In practice, the number of lags considered in estimating a FIGARCH model has to be truncated. We used lag truncation at 1,000 steps.[1] Because of the time needed for FIGARCH estimation, we only consider FIGARCH (1,d,1). Both GARCH and FIGARCH are estimated via the standard MLE procedures.

Since FIGARCH adopts the ARFIMA approach for modelling the dynamics of conditional volatility, one may ask whether one could not also use the original ARFIMA model as a possible data generating mechanism for financial volatility. The general ARFIMA model reads:

(4) $\Phi(L)\,(1-L)^d\,y_t = \Theta(L)\,\eta_t$

with $\Phi(L)$ and $\Theta(L)$ the AR and MA polynomials, respectively, and *d* the parameter of fractional differentiation. In our present application, y_t is given by squared residuals after filtering of linear dependence according to eq. (1). Like with GARCH and FIGARCH, we restrict ourselves to a maximum of one autoregressive and one MA term (i.e., $p \leq 1$ and $q \leq 1$). In contrast to FIGARCH, we also tried somewhat more parsimonious variants of the model allowing for $p = 0$ or $q = 0$ (this was possible because the computational burden for ARFIMA estimation is only a fraction of that necessary to obtain FIGARCH estimates). However, the specification $p = q = 1$ was almost always preferred. Estimation has first been tried via Fox and Taqqu's frequency domain maximum likelihood approach. However, when estimating the whole set of the parameters in this way, preliminary analysis of a smaller sample of time series showed extremely volatile and often very extreme results. We, therefore, resorted to estimating the fractional differentiation parameter via the Geweke and Porter-Hudak (1983) periodogram regression and, then, estimated the remaining parameters via the method of Fox and Taqqu assuming lag polynomials with roots strictly greater than 1 in modulus (which, in fact, seems to be the most popular method for estimating ARFIMA models in applied work).

Finally, the fourth model is the multi-fractal model of asset returns (Calvet and Fisher, 2001, 2002) in the version of Lux (2003). Essentially this is a stochastic volatility model whose volatility process, denoted by θ_t, is characterized by a multiplicatively connected hierarchy of random variables of various mean life-time:

(5) $\theta_t = 2^k \prod_{i=1}^{k} m_t^{(i)}$

The key distinguishing feature of this model is that the renewal of volatility components follows a stochastic process with different probabilities depending on their rank in the hierarchy of multipliers. Here we assume that Prob (new $m_t^{(i)}$) = $2^{-(k-i)}$ and that replacement of

[1] We also tried a moving lag length using all available past data plus 1,000 presample entries following the recommendation given in Chung (2002). However, this computationally even more burdensome practice produced virtually the same results.

a multiplier j implies simultaneous renewal of all i > j as well. This relatively simple and parsimonious construction allows for a variety of active components in current instantaneous volatility with mean live-time extending from $2^{-(k-1)}$ days for the first multiplier down to $2^{-(k-k)}$ = 1 day for the k-th component. Like in Calvet and Fisher (2002) and Lux (2003), the multipliers are drawn from a Lognormal distribution $m_t^{(i)} \sim LN\left(-\lambda \ln(2), s^2 \ln(2)^2\right)$ in which the second parameter *s* can be determined by the normalization of the mean value of each component to 0.5, $E[m_t^{(i)}] = 0.5$. Returns are obtained as a subordinate process with a multi-fractal instantaneous volatility according to (5):

(6) $x_t = \sqrt{\theta_t} \cdot \sigma_u \cdot u_t$

with $u_t \sim N(0, 1)$ and σ_u a stock-specific scale parameter.

Again, returns have been corrected for a constant mean and first-order serial dependence prior to estimation of this volatility model. For both estimation of the Lognormal multi-fractal model and its use for forecasting purposes, we follow Lux (2003) by implementing the GMM estimator devised in this paper with the same moment conditions (log increments at various lags together with their second moments).

In order to derive forecasts of future volatility (future squared returns) from the above models, different algorithms have to be used. While it is possible to explicitly derive conditional expectations for GARCH and FIGARCH models which, then, give the most efficient forecasts, this is not possible for MF and ARFIMA. In the later cases, we, therefore, resort to best linear forecasts (cf. Brockwell and Davis (1991, chap. 5) on the base of the autocovariances of these models for which closed form solutions can be obtained (cf. Beran, 1994, chap 8 for the ARFIMA process, and Lux, 2003 for the Lognormal MF model).[2] It is interesting to note that due to the use of moments of log increments in the GMM approach to estimation of the multifractal model, the scale parameter σ_u drops out from the moment conditions and cannot be estimated together with λ. However, this is not a drawback for forecasting purposes as any scale parameter would appear in both the numerator and the denominator of the coefficients of the best linear forecasts and would, therefore, drop out anyway. The MF model is, therefore, fully defined by the Lognormal parameter λ and the number of cascade steps, k.

2.2. Estimated Models

The parameter estimates of the GARCH, FIGARCH, ARFIMA and MF models are exhibited in *Tables 1* to *4*. From the roughly 1,200 stocks represented in the data base we have selected two subsamples: one with a random sample of 100 firms and another with the 100 companies with largest average trading volume. The appendix lists the names and length of the available time horizon (typically from 1975 to 2001 except for firms that were liquidated over the nineties). The length of the time series used for in-sample estimation of the

---

[2] For the simpler Binomial MF model of Calvet and Fisher (2001, 2003) explicit conditional expectations can be derived. However, a comparison between both forecasting algorithms (research in progress) shows practically no difference in performance.

parameters of the various models has been restricted to the 10 year period from 1975 to the end of 1984. Our main aim in restricting the in-sample period to roughly 40 percent of the data was to leave a relatively large sample for assessment of the forecasting quality of our models which then could be investigated over more than 15 years. Assuming a stationary volatility process according to one of our models, one may argue that ten years of data should be sufficient to estimate the model parameters reliably enough.

Table 1: GARCH parameter estimates

Random sample						
	mean	std	min	max	GARCH preferred	
ω	0.376	0.316	0.002	1.860	AIC	BIC
α_1	0.801	0.160	0.000	0.990	19	33
β_1	0.122	0.137	0.000	0.999		
$\alpha_1+\beta_1$	0.923	0.083	0.572	0.999		
Large volume						
	mean	std	min	max	GARCH preferred	
ω	0.345	0.320	0.004	1.978	AIC	BIC
α_1	0.769	0.165	0.000	0.983	16	22
β_1	0.151	0.135	0.016	0.999		
$\alpha_1+\beta_1$	0.920	0.081	0.474	1.000		

When inspecting the distribution of parameter estimates (whose mean, standard deviation, minimum and maximum across the pertinent subsamples are given in *Tables 1* to *4*), one remarks a relatively large variability. For example, both the parameters α_1 and β_1 of the GARCH model as well as the parameter of fractal differencing *d* in the FIGARCH model have values spread over the entirety of their admissible range [0,1]. The same variability applies to the ARFIMA's *d* although in the later case, we have not restricted the range of admissible vales to $d < 1$.[3] *Table* 1 also indicates how often the GARCH would be preferred over FIGARCH on the base of the Akaike and Schwartz information criteria (AIC and BIC). As it turns out, FIGARCH is preferred by about two out of three to four out of five cases and more so under AIC. This squares with the usual observation that BIC favours more parsimonious models. *Table 3* also reports the order of the ARFIMA models (p,d,q) with $p \in \{0,1\}$ and $q \in \{0,1\}$ estimated by the AIC criterion. The (1,d,1) model is the preferred one with only one exception in the 'large volume' sample (results with BIC are only marginally different).

---

[3] There is hardly any other choice than allowing for unrestricted d in the ARFIMA case since we cannot guarantee that the Geweke and Porter-Hudak estimates would fall into any prescribed interval. In contrast, the usual MLE estimation procedure for FIGARCH presupposes $0 \leq d < 1$.

Table 2: FIGARCH parameter estimates

	Random sample				Large volume			
	mean	std	min	max	mean	std	min	max
ω	0.503	0.395	0.002	1.795	0.405	0.336	0.023	1.868
β_1	0.546	0.304	-0.499	0.986	0.452	0.320	-0.535	0.922
φ_1	0.401	0.335	-0.621	0.959	0.340	0.336	-0.621	0.874
d	0.309	0.213	0.001	0.999	0.315	0.166	0.001	0.999

Table 3: ARFIMA parameter estimates: Volatility

Random sample							
Chosen models based on AIC				Estimate of d			
(1,d,1)	(1,d,0)	(0,d,1)	(0,d,0)	mean	std	min	max
100	0	0	0	0.218	0.153	0.001	0.975
Large volume							
Chosen models based on AIC				Estimate of d			
(1,d,1)	(1,d,0)	(0,d,1)	(0,d,0)	mean	std	min	max
99	0	1	0	0.254	0.143	0.002	0.992

Lastly, turning to our parameter estimates for the MF model, we again see a large variation of parameter values. Note that the lognormal distribution parameter λ is restricted to the open half line $[1,\infty)$. Estimates $\lambda = 1$ make the volatility cascade collapse to a constant value which leads to the benchmark case of Normally distributed returns. The mean values of the number of cascade steps *k* are about 12 and 14, for the 'random sample' and 'large volume' cases, respectively. It might be noted that, unlike ARFIMA and FIGARCH, the MF model is not a 'true' long-memory model, but only mimics hyperbolic decline of the autocorrelation function over about 2^k time lags after which one encounters an exponential drop-off of the ACF. Note that our average estimates of *k*, therefore, amount to slow decline of memory over up to 16,000 time steps – much more than used for estimating the model. The maximum estimate $k = 19$ even amounts to hyperbolic decline of the autocorrelation function over roughly half a million days, i.e. 2,000 years of daily trading. It is, therefore, obvious that the deviation of the MF model from 'true' long memory can be arbitrarily small. On the other hand, the minimum $k = 2$ has a range of hyperbolic scaling of just 4 days which makes it a rather clear-cut short-memory model.

Table 4: Multi-fractal parameter estimates: volatility

Random sample							
Estimate of λ				Estimate of k			
mean	std	min	max	mean	std	min	max
1.591	0.442	1.000	4.221	14.320	2.093	2.000	19.000
Large volume							
Estimate of λ				Estimate of k			
mean	std	min	max	mean	std	min	max
1.315	0.297	1.000	2.590	12.060	3.104	2.000	17.000

2.3 Forecasting Performance

Now turn to the results of our horse race for forecasting volatility: our estimated models have been tested out-of-sample for the 16-year period 1986 to 2001. Forecasting horizons start at the daily level and proceed via 5 day and 10 day forecasts up to 100 days ahead. Note that we have used only one set of parameter estimates and have not re-estimated the parameters within the out-of-sample period. The reason for not using rolling estimates is the computational burden of the FIGARCH model – with the other models it would have been feasible. We have also looked at the performance in subsamples (1986-1990,1991-1995, and 1996-2001), but to our surprise found no remarkable differences. As these periods cover quite diverse financial and economic conditions in Japan (including the stock market bubble, its crash and the subsequent stagnation) the homogeneity of the results speaks in favour of very regular *structure* in the volatility dynamics despite large changes in the *level* of volatility over time.

In order to compare the performance of the four candidate models, we apply the traditional concepts of mean squared error (MSE) and mean absolute error (MAE). However, since we want to have a meaningful measure allowing to compare the performance across stocks we have to standardize these statistics. We do so by reporting *relative* MSE and MAE obtained after division by the pertinent mean squared error and mean absolute error of the naïve predictor using historical volatility (i.e., the sample mean of squared returns over the period 1975 to 1984). In order not to compound errors in the mean equations and in the volatility dynamics, we also first filter out linear dependence analogously to eq. (1) and compute the naïve MSE and MAE from the squared residuals.

Our criteria for comparing predictive accuracy are, thus:

$$\text{(7) relative MSE} = \frac{1}{N}\sum_{t=1}^{N}(h_{t,j}-\varepsilon_t^2)^2 \Big/ \frac{1}{N}\sum_{t=1}^{N}(h_{t,n}-\varepsilon_t^2)^2 \, ,$$

$$\text{(8) relative MAE} = \frac{1}{N}\sum_{t=1}^{N}\left|(h_{t,j}-\varepsilon_t^2)\right| \Big/ \frac{1}{N}\sum_{t=1}^{N}\left|(h_{t,n}-\varepsilon_t^2)\right| \, ;$$

with $t = 1,\ldots, N$ the out-of-sample observations, $j = \{$ GARCH, FIGARCH, ARFIMA, MF $\}$ the estimates from the candidate time series models, the subscript n denoting the naïve predictions using historical volatility and ε_t the residuals obtained after linear filtering of returns (using the in-sample means and first-order autocorrelations).

Table 5 compares the average relative MSEs and MAEs of our four models for the ‘random sample’ of 100 stocks. Since results obtained for the ‘large volume’ cases are very similar, we have not reproduced them here for the sake of brevity, but are prepared to supply them upon request.

The winner in terms of average MSE reduction is the ARFIMA model (which so far has seldomly been considered as a model of volatility dynamics) followed by MF and FIGARCH. Interestingly, GARCH performs worse than all other models even over short horizons. The average improvement compared to naive forecasts are in the range of up to about 6 percent at daily horizons, 3 percent at ten days and still 1.5 percent at 100 day horizons.

To provide more details, *Fig*. 2 shows box plots of the distribution of MSEs and MAEs over all 100 stocks, for all methods and time horizons. A glance at the range of results for the different methods reveals some interesting tendencies. It particularly shows that for all long-memory models the inter-quartile range is below the benchmark of one for all time horizons. In contrast to GARCH, we, therefore, do find an improvement against the naïve prediction in the *majority* of cases with these models so that mean values above 1 in *Table 5* are due to very large entries in the upper end (the median is always < 1 for these methods). At the lower end, we find that for the one day horizon, MSE can be reduced against the naïve forecast by as much as 18 percent (GARCH, ARFIMA and FIGARCH and slightly less for MF) and still 7.5 percent improvement can be found at 100 day horizons for one particular stock (here GARCH seems to dominate slightly).

In terms of the upper end (the worst prediction within the sample), MF is best with a maximum MSE that *never* rises above unity (i.e., that is never worse than that of the naive forecasts). Note that it, therefore, can be described as the *least dangerous* method. This is in contrast to all other models with which the user always appears to face the danger of forecasts, that are worse than the most naive ones (i.e. max > 1). Note in particular how 'dangerous' GARCH and FIGARCH forecasts can become! Lastly, it appears noteworthy that ARFIRMA and MF have very small variability of their performance which also decreases with time horizon, while the fluctuations across assets in the (FI)GARCH models rather show a tendency for increasing dispersion at long horizons.

Table 5: Forecasting Volatility

	relative MSE				relative MAE			
horizon	GARCH	FIGARCH	ARFIMA	MF	GARCH	FIGARCH	ARFIMA	MF
1	0.954	0.947	**0.935**	0.944	1.060	1.084	1.045	**1.011**
5	0.985	0.964	**0.960**	0.971	1.080	1.106	1.055	**1.020**
10	1.000	0.975	**0.968**	0.978	1.092	1.118	1.058	**1.021**
20	1.019	0.985	**0.974**	0.982	1.114	1.136	1.057	**1.021**
30	1.043	0.997	**0.977**	0.985	1.138	1.153	1.056	**1.021**
40	1.071	1.008	**0.979**	0.987	1.162	1.168	1.054	**1.020**
50	1.106	1.021	**0.981**	0.988	1.187	1.184	1.053	**1.019**
60	1.145	1.036	**0.981**	0.988	1.211	1.199	1.051	**1.019**
70	1.192	1.053	**0.983**	0.989	1.237	1.214	1.050	**1.018**
80	1.243	1.071	**0.983**	0.989	1.261	1.229	1.049	**1.018**
90	1.300	1.089	**0.983**	0.990	1.286	1.243	1.048	**1.017**
100	1.370	1.110	**0.984**	0.990	1.311	1.258	1.047	**1.017**

Note: the 'winners' under each criterion are marked by bold numbers.

With respect to MAE, the multi-fractal model is the winner in all categories, over all time horizons. However, as a grain of salt, average performance of all models is worse than that of naive forecasts. On the other hand, the largest reduction of MAE achieved is as much as 24 % (1 day), 20 percent (10 days) and still 13 percent (100 days). Otherwise, results are comparable to those for the MSE criterion with a narrow range of entries for MF and a wide variation for FIGARCH and GARCH. Note also that MF comes closest to at least being 'neutral' under this criterion while all other methods have the inter-quartile range above the benchmark value of 1 and, therefore, lead to a deterioration against naïve forecasts in the majority of cases.

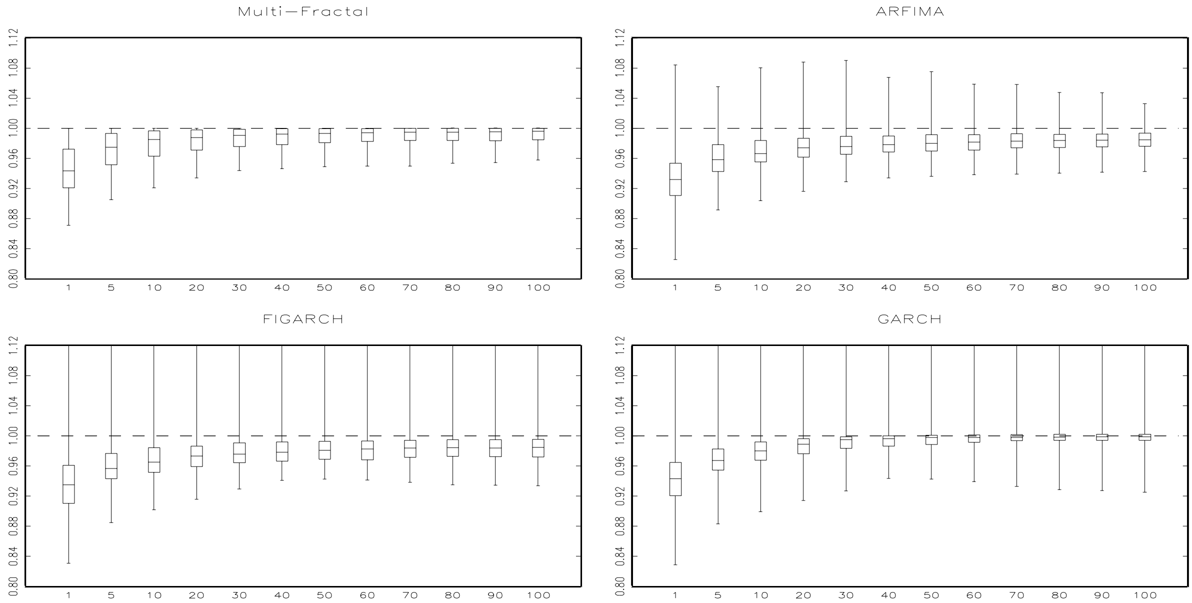

Fig. 2a: Distribution of MSEs of volatility predictions on the base of individual parameter estimates. The boxes show the median of the distribution surrounded by a box that spans the centre half of the data set (the inter-quartile range). The whiskers give the full range spanned by all 100 cases. For better comparability, we have chosen the same scale for all four box plots. The plots for the FIGARCH and GARCH results, therefore, do not show their respective maximum MSE which extends from 1.42 at lag 1 to 7.75 at lag 100 for FIGARCH (1.66 at lag 1 to 21.60 at lag 100 for GARCH).

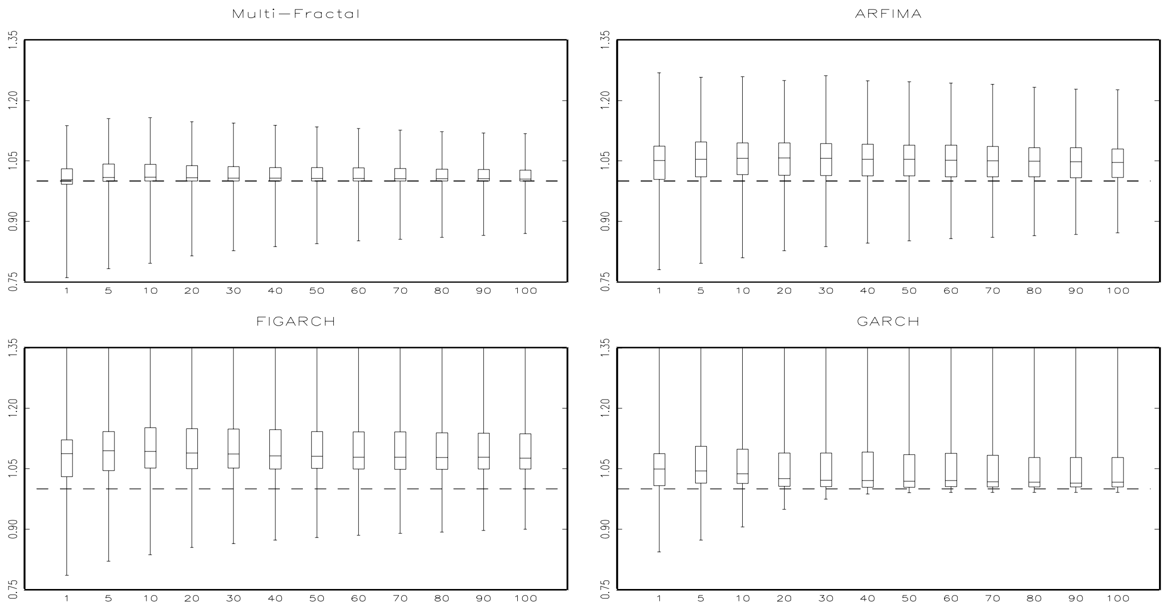

Fig. 2b: Distribution of MAEs of volatility predictions on the base of individual parameter estimates. For better comparability, we have chosen the same scale for all four box plots. The plots for the FIGARCH and GARCH results, therefore, do not show their respective maximum MAE which extends from 1.47 at lag 1 to 8.13 at lag 100 for FIGARCH (1.60 at lag 1 to 11.79 at lag 100 for GARCH).

A typical question arising in comparative studies of alternative predictors is whether the models under investigation use different information or not. The interesting consequence is that combinations of forecasts could improve results if the various models would not rely on the same information, whereas no such improvement appears feasible if their differences in performance are explained by different success in exploitation of the same underlying information. Typically one would use encompassing tests (Chong and Hendry, 1986) in order to shed light on this issue. However, our large sample of stocks renders this approach somewhat unpractical. Instead, we explore this question by computing the rank correlation of the forecasting success across all assets for each pair of methods. A high entry would suggest that two methods use virtually the same information so that the difference in their relative MSEs and MAEs is mainly to be explained by difference in the accuracy of the conditional expectations. Low rank correlation, on the other hand, might suggest room for improvement via forecast combination.

Table 6: Rank Correlations of Volatility Predictions Across Assets

Random sample: relative RMSE						
lead	GARCH-FIGARCH	GARCH - ARFIMA	GARCH-MF	FIGARCH-ARFIMA	FIGARCH - MF	ARFIMA-MF
1	0.896	0.610	0.423	0.558	0.447	0.442
5	0.843	0.612	0.323	0.701	0.462	0.373
10	0.732	0.512	0.224	0.699	0.430	0.352
20	0.578	0.430	0.170	0.697	0.423	0.345
30	0.489	0.305	0.082	0.690	0.393	0.327
40	0.404	0.219	0.083	0.686	0.392	0.320
50	0.372	0.137	0.124	0.653	0.389	0.316
60	0.348	0.110	0.127	0.651	0.396	0.318
70	0.340	0.076	0.125	0.624	0.386	0.309
80	0.335	0.053	0.145	0.612	0.385	0.320
90	0.328	0.046	0.154	0.609	0.403	0.327
100	0.335	0.028	0.167	0.604	0.399	0.326
Random sample: relative MAE						
lead	GARCH-FIGARCH	GARCH - ARFIMA	GARCH-MF	FIGARCH-ARFIMA	FIGARCH - MF	ARFIMA-MF
1	0.767	0.664	0.471	0.667	0.510	0.462
5	0.737	0.555	0.326	0.565	0.390	0.374
10	0.686	0.494	0.275	0.537	0.370	0.348
20	0.622	0.378	0.189	0.520	0.330	0.321
30	0.574	0.289	0.152	0.495	0.303	0.284
40	0.551	0.235	0.140	0.487	0.280	0.276
50	0.527	0.173	0.097	0.462	0.265	0.256
60	0.518	0.121	0.073	0.446	0.258	0.251
70	0.510	0.093	0.054	0.434	0.257	0.248
80	0.500	0.082	0.050	0.425	0.257	0.243
90	0.500	0.068	0.041	0.419	0.254	0.241
100	0.497	0.065	0.038	0.415	0.247	0.244

Note: At a significance level of 95 percent, the null hypothesis of no correlation in the performance of different methods would have to be rejected for absolute entries beyond 0.197 (=1.96 $\sqrt{(n-1)}$) with n = 100 in our case).

Table 6 gives the rank correlation across assets for all pairs of two methods for both relative MSE and relative MAE. If all methods would have the same ranking of MSEs and MAEs across assets, rank correlations would be 1. This is not the case: although a relatively large rank correlation exists at small horizons, different methods are more or less successful in predicting the volatility of individual assets. This implies that they are not simply using the same information more or less efficient, but that they might perform differently on different assets. Combination of forecasts, therefore, might still improve the overall results. Furthermore, the highest correlations exist between FIGARCH and GARCH at small forecasting horizons and between FIGARCH and ARFIMA at longer horizons pointing to the built-in similarities in their behaviour for short and long time horizons, respectively.

3. Forecasting Volume

3.1. Models

The empirical finance literature has mainly concentrated on trying to predict returns and volatility, but has hardly paid any attention to volume: a systematic search has brought about one single entry, Kaastra and Boyd (1995). These authors use neural networks and ARIMA models to forecast monthly futures trading volume for the Winnipeg Commodity Exchange. They emphasize the practical implications of volume predictions for the operation of the exchange. Besides its importance for forecasting transaction fees and liquidity, we may add that volume forecasting is also interesting in view of the similarity of the time series properties of both volatility and volume. Given the evidence on similar long term dependence in both series (Bollerslev and Jubinski, 1999; Ray and Tsay, 2000) it seems interesting to explore whether models with this feature are similarly capable of predicting both future volume and volatility.

We, therefore, continue our study by also using the volume entries in our data base to investigate the forecastability of transaction volume. To enhance comparability with the results obtained for volatility, we use again both the stocks represented in the 'random sample' and the selection based on the average volume itself. Since results are again quite similar, we only exhibit those for the randomly selected stocks and provide additional results for the high volume cases upon request. As before, we estimate models on the base of the eleven year period 1975 through 1985 and test the forecasting performance of the models for the remaining fifteen years (sometimes less) 1986 to 2001 thus allowing an assessment of their success over a relatively long time horizon. When investigating volume data of stock exchanges, researchers typically find that these are non-stationary and have to be detrended first before they can be used to shed light on volatility and return dynamics. Interestingly, considering the 27 year period from 1975 to 2001 as a whole, trends in trading volume are practically non existent in the Japanese market. This is readily apparent form the quite typical behaviour of the volume of the Nippon Suisan Kaisha share exhibited in *Fig*. 1: while long subsets of the data from 1975 to about 1990 would have given rise to apparently positive time dependence, the later development suggests that the increase of volume in second half of the eighties should be interpreted as an intermittent episode rather than the signature of a secular trend. Given this absence of clear trends, we use the *raw volume data without any correction or detrending* in our subsequent forecasting exercise.

Because of the lack of applicability of the GARCH family, only three models have been estimated for the volume time series:

As a short-memory benchmark we estimate an ARMA(p,q) model: in order to see whether a moderate number of lags suffices to capture the time dependence in volume records, we select an ARMA(p,q) model for forecasting within the range $p \leq 5$ and $q \leq 5$ via maximum likelihood and use the Akaike criterion for selection of one of these alternatives. We deliberately chose AIC rather than the typically more parsimonious BIC criterion in order to allow for a sizable number of lags which could suffice for modelling the dynamics without having to resort to genuine long memory models. However, despite BIC's tendency towards more parsimonious models, results with respect to forecasting quality turned out to be similar.

Our second model is the fractionally integrated ARFIMA(p,d,q) model. Because of the higher computational burden and also because longer lags should be captured by the fractional differentiation term, we restrict ourselves again to a maximum of one autoregressive and one MA term (i.e., $p \leq 1$ and $q \leq 1$). Estimation proceeds along the same lines as in the application of ARFIMA to volatility. For both ARMA and ARFIMA models, estimation is restricted to lag polynomials with roots strictly greater than 1 in modulus. the ARFIMA models we allowed for non-stationary variants by estimating the ARFIMA model with differenced data when the initial GPH estimate of the fractional differencing parameter *d* exceeded the benchmark 0.5. Forecasting, then, is performed by integrating the forecasts of the differenced series.

Third, we also apply the multi-fractal cascade process as a model for volume. Since volume has a structure similar to volatility, we simply adopt the volatility cascade part depicted in eq. (5):

(9) $$\theta_t = 2^k \prod_{i=1}^{k} m_t^{(i)} ,$$

but skip the incremental Normal distribution introduced in eq. (6) which in the volatility model mainly serve to randomise the sign of returns. All that is needed to use this as a model of the volume dynamics is an additional scaling factor to capture the different size of mean volume in each stock. Hence, volume can be written as $vol_t = \theta_t \cdot \chi_i$ where χ_i is the scaling factor for asset i. However, one should note that we neither need the scale parameter for estimating the other parameters nor in constructing forecasts so that χ_i is, in fact, only introduced to organize our discussion of the model. Since estimation is again based on GMM with log increments, this parameter drops out and only the location parameter of the Lognormal multipliers, λ, has to be estimated. And χ_i is also not needed in our forecasting exercise as any scale parameter would appear in both the numerator and the denominator of the coefficients of the best linear forecasts and would, therefore, drop out anyway. Again, the MF model is, therefore, fully defined by the Lognormal parameter λ and the number of cascade steps, k.

Since our short-memory benchmark is now an ARMA(p,q) model, we also do not subject the data to filtering out linear dependency. The multi-fractal cascade, therefore, is applied to volume in its basic format (eq. 9) without any additional adjustments. An advantage of MF against ARMA and ARFIMA models might be seen in the fact that by its very definition, it allows for positive entries only while negative realizations cannot be excluded in the two alternative models. We, therefore, used zero as the lower bound for our forecasts from ARMA and ARFIMA models (which, however, was hardly ever a binding constraint).

In our initial tests, we also estimated ARIMA(p,1,q) models. However, since it turned out that their forecasting performance was almost always far worse than that of the alternative models, we dropped them from our final design of this forecasting exercise.

3.2. Parameter Estimates

Now turn to the estimation results: *Table 7*, first, shows that ARMA estimation tends to favour models with many parameters at least under the AIC criterion. Comparing the AIC and SIC model selection criteria for the preferred ARMA and ARFIMA models, we see that AIC would prefer the ARMA over the ARFIMA specification in 98 out of 100 cases for both the large volume and random sample. Hence, the long-term dependence seems not to be able to compensate for the admission of more AR and MA components in our ARMA design. The SIC, however, produces fewer cases of preferred ARMA models confirming the well-known finding that it typically favours more parsimonious models than AIC.

Table 7: ARMA parameter estimates

Random sample							
Chosen models						ARMA preferred	
(5,5)	(5,4)	(5,3)	(4,5)	(3,5)	other	AIC	SIC
35	14	6	14	10	21	98	58
Large volume							
Chosen models						ARMA preferred	
(5,5)	(5,4)	(5,3)	(4,5)	(3,5)	other	AIC	SIC
26	19	10	16	6	23	98	58

Table 8 shows that the ARFIMA model produces higher dispersion of its key parameter *d* for the random sample albeit with a lower average *d* than for the large volume firms. The preferred type of model is overwhelmingly the (1,d,1) variant modulating the prevalent long-term dependence via additional AR and MA components. Due to the wide variability of the AR and MA components, their statistics are not shown but are available upon request. Comparison with volatility (*Table 3*) reveals no clear tendency of the differences: while mean estimates of *d* are higher with volume, the minima and maxima across samples are lower.

Table 8: ARFIMA parameter estimates: Volume

Random sample							
Chosen models				Estimate of d			
(1,d,1)	(1,d,0)	(0,d,1)	(0,d,0)	mean	std	min	max
97	1	2	0	0.291	0.135	0.002	0.671
Large volume							
Chosen models				Estimate of d			
(1,d,1)	(1,d,0)	(0,d,1)	(0,d,0)	mean	std	min	max
91	5	4	0	0.344	0.121	0.074	0.639

Table 9 exhibits information about the parameters of the estimated multi-fractal model. One can infer that the estimates of the key parameter λ are all within the interval between 1.00 and about 1.2 (1.00 being the lower limit for this parameter). Estimated λ's are somewhat higher on average for the random sample of stocks signalling larger bursts of activity. This might be explained by a larger increase of trading volume during the bubble for the average firm compared to large firms which already had relatively high trading volume before the bubble episode. Besides that, we find a higher dispersion of both the λ and k estimates for the random sample compared to the large volume firms which shows a higher degree of homogeneity of the time series characteristics of the later sample. Compared with *Table* 4, we see that both the estimated λs as well as the number of cascade steps are smaller on average for volume than for volatility.

Table 9: Multi-fractal parameter estimates: volume

Random sample							
Estimate of λ				Estimate of k			
mean	std	min	max	mean	std	min	max
1.118	0.039	1.000	1.221	9.250	2.153	2	13
Large volume							
Estimate of λ				Estimate of k			
mean	std	min	max	mean	std	min	max
1.089	0.027	1.050	1.195	8.950	1.617	6	12

3.3. Forecasting Results

Table 10 shows the forecasting performance of the ARMA, ARFIMA, and MF models over forecasting horizons of 1, 5, 10, 20 etc. up to 100 days for the out-of-sample period 1986 to 2001. We resort again to the criteria of relative mean squared error (MSE) and relative mean absolute error (MAE) for our assessment of the forecasting performance. The table shows the mean relative MSEs and MAEs over the 100 time series from the random sample of stocks (results for the 100 stocks with the highest average trading volume are again quite similar and can be obtained upon request). The results are perplexing: in both categories, the multi-fractal model has lowest average MSEs and MAEs over most time horizons. Furthermore, these means are all smaller than in the case of volatility signalling a sizable average gain in forecasting performance against the naïve model (i.e., the in-sample mean value of the time series). Roughly, MF achieves an average improvement of 45 percent (MSE) or 29 percent (MAE) for one-day horizons and even over a forecasting horizon of 100 days has a performance that is by about 4 -5 percent better in both criteria than the naïve model. The ARFIMA forecasts mostly reach second rank with averages only slightly above those of MF, while ARMA yields much poorer results (implying a deterioration against the naïve forecasts for all but the 1 day horizon).

As can be seen from *Fig. 3*, the MF model also has again the smallest standard deviation of its forecast errors showing that the success of this method is more uniform than that of ARFIMA and ARMA models. While ARFIMA is comparable to MF in its average performance (as indicated by its inner-quartile ranges), it has much higher cases of 'failure' (the upper and of the whiskers). MF's maxima, in contrast, do only very slightly exceed the benchmark value of 1, so that here the danger of getting worse forecasts than with the naive model is almost nonexistent. As with GARCH and FIGARCH in the case of volatility forecasting, ARMA models of volume produce very unsatisfactory results. Not only does one face the danger of extremely poor entries (with MSE up to 146 times that of the naïve model in one case), but rather the whole ensemble of 100 forecasting exercises performs quite poorly. ARMA only produces an improvement against the naïve forecasts for the one day horizon and does hugely worse thereafter. It is particularly astonishing that even for the one day horizon, ARMA is dominated by both ARFIMA and MF throughout and that it performs the worst at the five day horizon although it typically uses 5 lags in either the MA or AR component or both.[4]

The most astonishing feature is, however, the success of the MF model which even for small horizons is better than its competitors – although one estimates only two parameters and unlike in the AR(FI)MA classes there are no parameters available in this model for fine-tuning of short-term dependence. Nevertheless, the MF model mostly produces better short-term forecasts than the naïve benchmark prediction and even in its 'bad' cases has relative MSEs and MAEs only slightly above one while the ARFIMA and particularly the ARMA models can be far off the mark.

These results are also interesting from the perspective of the theoretical literature on forecasting on the base of ARMA and ARFIMA models. Most perplexingly, the results exhibited for ARFIMA (and MF) in Table 10 are close to (if not better than) what one could expect to obtain for an ARFIMA process with *known* parameter values. Beran (1944, chap. 4) computes MSE improvements over the (known) unconditional variance from best linear forecasts with an infinite number of past observations. In his Table 8.8 we find for d = 0.1 und d = 0.4 improvements by 1.91 % and 48.82 % (one step ahead), 0.22 % and 27.06 % (ten steps), and 0.03 % and 14.75 % (hundred steps). A rough interpolation between these extremes with our mean estimates of 0.29 from Table 8, in fact, indicates that our volume forecasts are pretty much in line with what could be expected with a 'true' underlying ARFIMA model even without accounting for parameter uncertainty.

From this perspective, however, the poor performance of the ARMA class is surprising as several papers show that suitably adapted ARMA models can produce forecasts comparable to that of 'true' underlying ARFIMA models (Basak et al., 2001; Man, 2003). This divergence might be explained by different factors: on the one hand, estimated fractional differentiation parameters of our data are relatively high so that it is hard to cover the persistency of the data by short-memory models. In fact, several papers have shown that the approximation of long-memory models by ARMA structures works best for small values of d and becomes less satisfactory for strongly dependent processes (Brodsky and Hurvich, 1999; Crato and Ray, 1996). Furthermore, our ARMA models have been chosen by the usual AIC (or BIC) criteria and, therefore, are not those optimally adapted for forecasting an assumed underlying long-memory process. In any case, the results illustrates that the choice between short-memory and long-memory processes can crucially affect forecasting performance (even over short horizons).

---

[4] ARFIMA performs worse for the large volume than for the random sample but inspection shows that the difference is actually only due to one extreme outlier in the large volume sample. Eliminating this entry, results for large volume become almost the same as for the random sample.

Table 10: Forecasting Volume

	relative MSE			relative MAE		
horizon	ARMA	ARFIMA	MF	MF	ARFIMA	MF
1	0.822	0.560	**0.556**	ARMA	0.713	**0.707**
5	2.784	0.827	**0.815**	0.938	0.890	**0.873**
10	1.976	0.880	**0.869**	1.287	0.932	**0.911**
20	1.549	0.911	**0.903**	1.260	0.962	**0.936**
30	1.340	0.927	**0.920**	1.236	0.978	**0.949**
40	1.294	0.934	**0.927**	1.184	0.984	**0.953**
50	1.142	0.938	**0.934**	1.171	0.990	**0.957**
60	1.089	0.941	**0.938**	1.134	0.991	**0.958**
70	1.069	0.945	**0.943**	1.112	0.995	**0.960**
80	1.061	**0.944**	0.945	1.096	0.995	**0.961**
90	1.064	**0.945**	0.948	1.087	0.996	**0.962**
100	1.057	**0.947**	0.949	1.085	0.997	**0.962**

Note: the 'winners' under each criterion are marked by bold numbers.

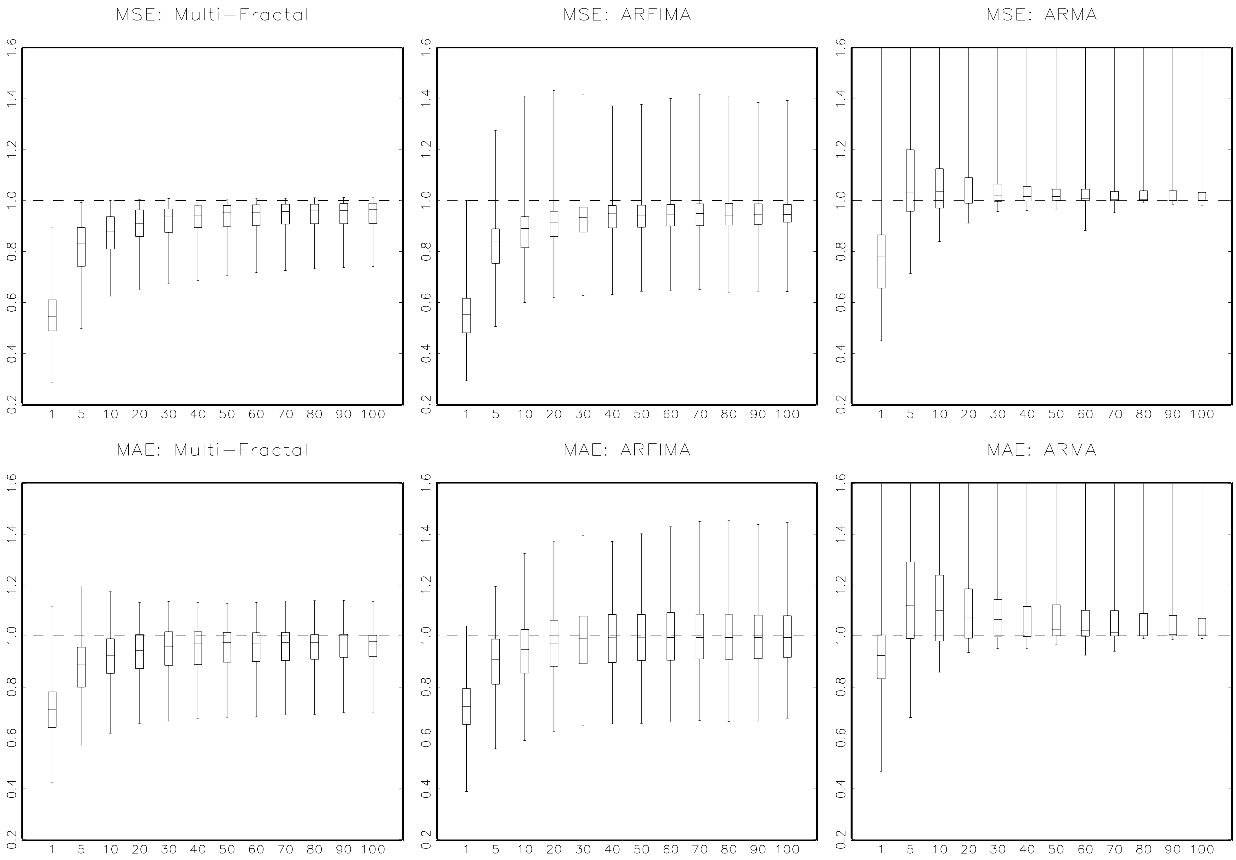

Fig. 3: Distribution of MSEs and MAEs of volume predictions on the base of individual parameter estimates. The boxes show the median of the distribution surrounded by a box that spans the centre half of the data set (the inter-quartile range). The whiskers give the full range spanned by all 100 cases. For better comparability, we have chosen the same scale for all four box plots. The plots for the ARMA results, therefore, do not show their respective maximum values which extend from 4.25 at lag 1 to 2.90 at lag 100 for MSEs (2.29 at lag 1 to 2.47 at lag 100 for MAEs). Interestingly, the MSEs and MAEs of the estimated ARMA models exhibit an inverted U shape in most cases with maximum errors at the 5 day forecasting horizon (the maxima over all stocks at the 5 day horizon are 146.19 for MSE and 8.94 for MAE).

Again, we try to provide an overall assessment of the degree of complementary between methods. To this end, Spearman's coefficients of rank correlation are exhibited in *Table 11* for both the MSE and MAE values achieved at various forecasting horizons by the various time series models. Interestingly, the correlation between the success of the MF and ARFIMA models is highly significant over all forecasting horizons. This means that if MF produces a high (low) reduction of MSE and MAE against the naïve model, the same is also likely to happen for the ARFIMA model. Hence, exploitation of information from past variables is quite uniform with both models: cases in which one model is particularly good while the other performs very poorly occur relatively seldom. Mostly, good results obtained from MF coincide with relatively good forecasting performance of ARFIMA as well. In contrast, correlation between the long-memory models and the ARMA model is significant only for few lags and is uniformly smaller than the MF-ARFIMA correlation.

Table 11: Rank Correlations of Volume Forecasts Across Assets: Random Sample

	MSE			MAE		
lead	ARMA-ARFIMA	ARMA-MF	ARFIMA-MF	ARMA-ARFIMA	ARMA-MF	ARFIMA-MF
1	0.511	0.443	0.914	0.523	0.526	0.929
5	0.170	0.240	0.854	0.219	0.236	0.914
10	0.202	0.267	0.825	0.182	0.181	0.884
20	0.139	0.180	0.841	0.145	0.174	0.866
30	0.082	0.060	0.832	0.109	0.138	0.845
40	0.039	0.011	0.834	0.096	0.103	0.833
50	-0.020	-0.080	0.858	0.069	0.067	0.829
60	-0.026	-0.070	0.832	0.066	0.094	0.824
70	-0.074	-0.092	0.823	0.042	0.053	0.814
80	-0.097	-0.142	0.823	0.049	0.020	0.817
90	-0.099	-0.151	0.823	0.053	-0.012	0.812
100	-0.116	-0.164	0.803	0.036	-0.005	0.811

Note: At a significance level of 95 percent, the null hypothesis of no correlation in the performance of different methods would have to be rejected for absolute entries beyond 0.197 (= 1.96 √(n-1)) with n = 100 in our case).

4. Forecasting with Pooled Estimates

Inspection of parameter estimates and forecasting results for GARCH, FIGARCH, ARMA and ARFIMA models across stocks shows that the worst results are obtained with extreme parameter estimates. For example, in volatility forecasting GARCH and FIGARCH performance is worst for nearly integrated processes, i.e. $\alpha_1 + \beta_1 \approx 1$ in the GARCH and $d \approx 1$ in the FIGARCH model, respectively. Similarly, the performance of the ARMA models for volume often becomes extremely poor when one of the roots approaches one. For ARFIMA, we also encounter relatively poor results for high estimates of *d*. Interestingly, the problem of extreme failures in some individual stocks seems non-existent for the multi-fractal model which in this important sense appears to be much more robust – in both its application to volatility and volume – than all the more traditional models. Particularly striking is the observation that at least the MSE of volatility forecasts does never (in not one single case of

100 stocks and in not one single forecasting horizon) exceed that of the naïve estimator.[5] Interestingly, application of ARFIMA models to volatility seems to be much 'safer' than their use in forecasting of volume (the maximum deterioration in MSEs against the naive forecasts are about 10% for volatility against more than 50% for volume).

The big failures of some methods with some series could have quite different sources: first, if the underlying time series were, in fact, non-stationary or almost non-stationary, it might simply be that their degree of forecastibility is lower than for some other series. Along a similar line of argument, they might simply possess some large outliers (remember that we included the bubble period) or other particularities, which could have affected our forecasting results. Interestingly, our comparative investigation of alternative forecasts seems to allow to safely exclude this possibility: in all cases there had at least been one method whose forecasts did not perform too badly for exactly the same series (i.e., the MF model).

An alternative explanation would, therefore, have to look for the fault in the parameter estimates of the poorly performing models. The poor forecasts might then be attributed to the variability of parameter estimates with extreme failures being due to rather large random deviations between estimated and 'true' parameters.[6]

In order to see whether restricting the variability of parameter estimates allows us to avoid the defective results in some cases, we resorted to forecasting with the mean parameter estimates obtained across our 100 randomly selected stocks.

The average estimates are taken from *Tables 1* to *4* for the volatility models and from *Tables 7* to *9* for the models used to forecast volume. To account for varying scale of the fluctuations across stocks, the following adjustments have to be made:

For the GARCH models, we now forecast on the base of the average parameters $\bar{\alpha}_1 = 0.801$ and $\bar{\beta}_1 = 0.122$. Since the remaining parameter, ω, gives the scale of fluctuations (with the unconditional variance equal to $\frac{\omega}{1-\alpha_1-\beta_1}$) it would hardly be useful to average this coefficient across stocks. Instead we compute ω from the unconditional sample variance of each (linearly filtered) return series, $\hat{\sigma}^2$, using the average estimates of the remaining parameters: $\omega = \left(1 - \bar{\alpha}_1 - \bar{\beta}_1\right) \cdot \hat{\sigma}^2$. Alternatively, we used averages for the dynamic parameters α_1 and β_1 but kept the previous stock-specific estimate of ω which yielded practically identical results.

In the FIGARCH model, the mean volatility level (the conditional variance) is not determined. However, in practice one has to approximate the fractional difference operator on the RHS of eq. (3) by a finite approximation of its expansion. The chosen cut-off of the infinite sum, then, in fact guarantees existence of the unconditional variance which is given

---

5 Practically the same outcome applies when the 100 stocks with the highest trading volume are chosen rather than a random selection of stocks.

6 To be precise, we cannot really speak of 'true' parameters in a comparative study of various models of which none will be the 'true' data generating mechanism. One might, therefore, rather think of the 'true' parameters as those that best represent the particular class of models for a certain purpose (forecasting).

by $\frac{\omega}{\delta_d(1)}$ with $\delta_d(L) = (1-L)^d \approx \sum_{k=0}^{k_{max}} \delta_{d,k} \cdot L^k$ where the summation up to the cut-off k_{max} instead of the infinite theoretical sum leads to a non-vanishing $\delta_d(1) \neq 0$ (Chung, 2002). Taking this implication of the practical approach to FIGARCH modelling into account, we can fix the parameter ω similarly as for the GARCH model in order to capture the different scales of fluctuations for individual assets. Alternatively, we also tried the average ω (along with average values of α_1, β_1 and *d*) for all stocks which produced practically identical results).

For the multi-fractal model we simply took the mean estimate of the crucial parameter λ together with the mean of the number of cascade steps rounded to the nearest integer. Note that unlike in the GARCH and FIGARCH cases, for the multi-fractal as well as the ARFIMA models, the scale parameter is irrelevant as the weights in the best linear forecasts are scale-free (more precisely, the scale parameter appears in both the numerator and denominator and, therefore, does not affect the forecasts).

For those models where we allowed for flexible choice of the number of lags (ARMA and ARFIMA), the mean estimates were computed for the maximum number of lags. The coefficients of the average estimates are, then, the means over the one hundred individual samples with cases of more parsimonious models contributing a zero value for their missing coefficients.

The results of our exercise are quite striking: overall, we mostly see an improvement of forecasting performance when using average instead of individually optimised parameters. *Table 12* details our results for volatility: under the MSE criterion, we find improvements for all models under almost all perspectives. In particular, the mean MSE is always smaller than with the individual parameter estimates with the improvement being most pronounced for FIGARCH and GARCH whose performance was lacking behind ARFIMA and MF when using individual parameter estimates. As a result, the three long memory models are now head to head with MF slightly behind FIGARCH and ARFIMA. Again, for all lags (even the smallest ones) GARCH despite its improvement falls clearly behind the long memory models. What is more, a look at *Fig*. 4 shows that this better average performance does not come at the price of deterioration of the best cases (the lower part of the whiskers shows little variation between *Figs*. 2 and 4 and if anything the ‘best’ cases become even better in the case of pooled estimates). It, therefore, appears that the whole distribution of forecasting results seems to shift to the left. Overall, under the MSE criterion, averaging appears almost unambiguously superior as it not only improves forecasting performance in good cases but also appears to minimize the risk of poor predictions (all the maximum entries are now close to 1).

Table 12: Forecasting Volatility: Pooled Estimates

relative MSE					relative MAE			
horizon	GARCH	FIGARCH	ARFIMA	MF	GARCH	FIGARCH	ARFIMA	MF
1	0.926*	0.920*	**0.916***	0.919*	1.037*	1.083*	1.053*	**1.036**
5	0.966*	0.951*	**0.949***	0.952*	**1.039***	1.110*	1.068*	1.047
10	0.977*	0.959*	**0.958***	0.962*	**1.024***	1.118	1.070*	1.048
20	0.986*	**0.963***	0.965*	0.969*	**1.006***	1.123*	1.068*	1.045
30	0.994*	**0.968***	0.969*	0.974*	**1.002***	1.126*	1.066*	1.043
40	0.997*	**0.969***	0.971*	0.976*	**1.001***	1.126*	1.063*	1.041
50	0.999*	**0.971***	0.973*	0.978*	**1.000***	1.127*	1.062*	1.039
60	0.999*	**0.972***	0.975*	0.979*	**1.000***	1.128*	1.060*	1.038
70	1.000*	**0.973***	0.976*	0.980*	**1.000***	1.129*	1.059*	1.037
80	1.000*	**0.973***	0.977*	0.981*	**1.000***	1.129*	1.057*	1.035
90	1.000*	**0.974***	0.977*	0.982*	**1.000***	1.129*	1.056*	1.034
100	1.000*	**0.974***	0.978*	0.983*	**1.000***	1.129*	1.055*	1.033

Note: the 'winners' under each criterion are marked by bold numbers. The asterisks indicate an improvement in average MSE and MAE against the forecasts with individual parameter estimates in sec. 2. Note that pooled estimates lead to improvements for all models under the MSE criterion and for all but MF (and FIGARCH at 10 day horizon) under the MAE criterion.

Results are somewhat different under the MAE criterion: here we find a slight deterioration for MF, but again improvements for GARCH, FIGARCH and ARFIMA. The winner is the GARCH model while we had a clear dominance of the MF model for individually optimised estimates. However, with the transition form the formerly winning MF to GARCH and ARFIMA as the best performing alternatives in the pooled estimation exercise no real gain is achieved under the MAE criterion since the average GARCH prediction essentially coincides with the naïve forecast for horizons of 20 days and more. Therefore, the major gain would consist in a reduction of the relative MSE of the worst performing cases.

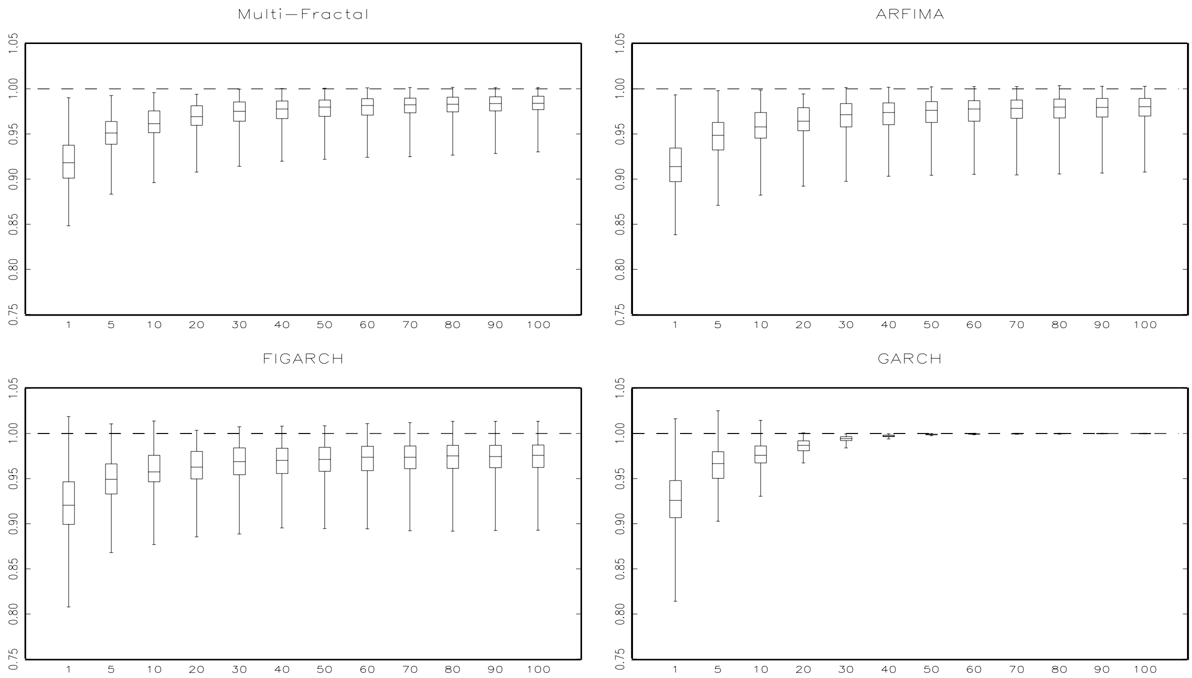

Fig. 4a: Distribution of MSEs of volatility predictions on the base of pooled parameter estimates. For the construction of the box plot, cf. the legend of *Fig*. 2. Apparently, the danger of poor predictions is dramatically reduced.

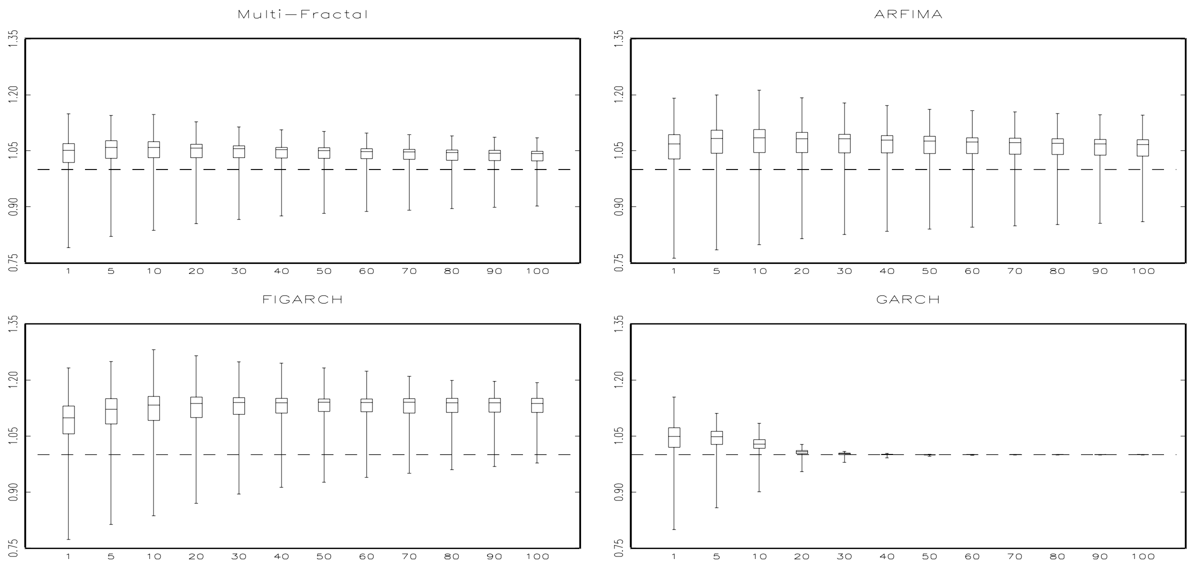

Fig. 4b: Distribution of MAEs of volatility predictions on the base of pooled parameter estimates. For the construction of the box plot, cf. the legend of *Fig*. 2.

Now turn to volume: again we find mostly improvements under the MSE criterion when replacing the individual parameter estimates by pooled estimates (*Table 13*). Improvements are more spectacular for ARFIMA and particularly so for ARMA, which had a relatively poor performance with individual estimates, while only slight positive and negative changes can be identified for the MF model. As can be seen from *Fig.* 5, the overall performance of MF and ARFIMA is now virtually the same for the MSE criterion, but it shows a certain advantage for MF under the MAE criterion. Similarly as with for volatility the most striking finding is the highly reduced danger of poor performance, particularly so for ARMA and ARFIMA models.

Table 13: Forecasting Volume: Pooled Estimates

	relative MSE			relative MAE		
horizon	ARMA	ARFIMA	MF	ARMA	ARFIMA	MF
1	0.647*	0.541*	**0.552***	0.792*	**0.704***	0.706*
5	0.978*	**0.803***	0.811*	0.991*	0.891	**0.873**
10	0.987*	**0.859***	0.865*	0.990*	0.936	**0.910***
20	0.996*	**0.892***	0.898*	0.994*	0.967	**0.932***
30	0.998*	**0.910***	0.917*	0.996*	0.984	**0.945***
40	0.999*	**0.917***	0.925*	0.997*	0.991	**0.948***
50	0.999*	**0.923***	0.933*	0.997*	0.997	**0.952***
60	0.999*	**0.924***	0.936*	0.998*	0.998	**0.951***
70	0.999*	**0.928***	0.942*	0.998*	1.001	**0.953***
80	0.999*	**0.930***	0.945	0.998*	1.002	**0.954***
90	0.999*	**0.931***	0.947*	0.998*	1.003	**0.955***
100	0.999*	**0.931***	0.950*	0.998*	1.002	**0.955***

Note: the ‘winners’ under each criterion are marked by bold numbers. The asterisks indicate an improvement in average MSE and MAE against the forecasts with individual parameter estimates in sec. 2. Note that pooled estimates lead to improvements for practically all models and horizons under the MSE criterion and for all but ARFIMA under the MAE criterion.

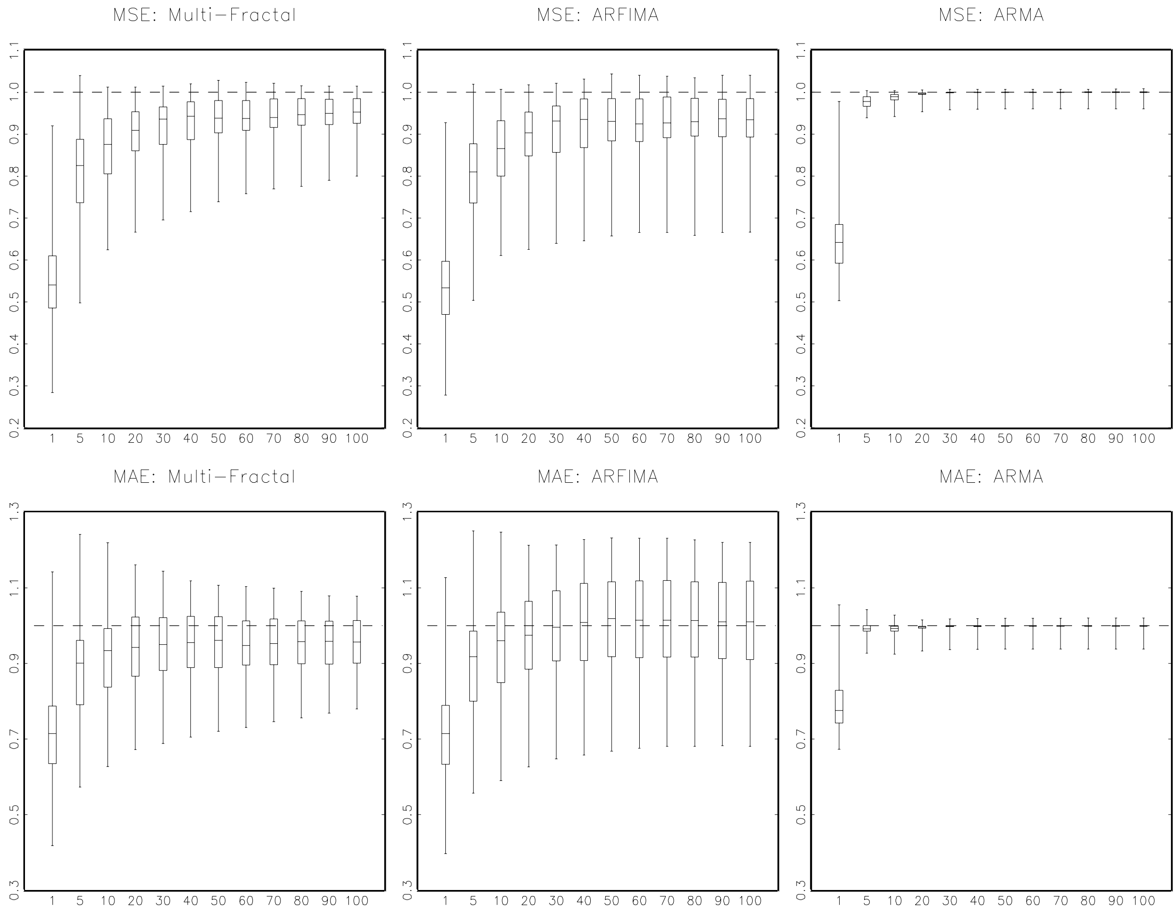

Fig. 5: Distribution of MSEs of MAEs of volume predictions on the base of pooled parameter estimates. For the construction of the box plot, cf. the legend of *Fig.* 2. Apparently, the danger of poor predictions is dramatically reduced.

Overall, it appears that using pooled estimates is more useful in improving predictive power for volatility models than for volume prediction, but for both volume and volatility it greatly reduces the danger of arriving at very poorly performing models. It is also instructive to compare the rank correlations between methods for pooled estimates (*Tables 14 and 15*) to those computed for our original estimates (*Tables 6* and *11*). In contrast to sec. 3 and 4, we now find a much higher correlation among the long memory models over all time horizons which even at a forecasting horizon of one-hundred days remains above 90 percent. This indicates that pooled estimates of different models exploit the same features of the data. Combination of forecasts would, then, surely not improve upon the performance of the best model. In contrast, rank correlation between long memory (FIGARCH, ARFIMA and MF) and short memory models (GARCH and ARMA) are decreasing much faster with increasing time horizon. While they remain somewhat significant for volatility models, they clearly become insignificant for the volume models.

Taking into account the better performance of the long-memory models, this result underscores that the data exhibit features, which are solely detected by models with long-term dependence. Since pooled GARCH and ARMA forecasts are mostly undistinguishable from naïve forecasts at long horizons, this finding speaks in favour of the "true" presence of long correlations in both the volume and volatility data.

Table 14: Rank Correlations Across Assets: Volatility Forecasts, Random Sample

	MSE					
lead	GARCH-FIGARCH	GARCH - ARFIMA	GARCH-MF	FIGARCH-ARFIMA	FIGARCH -MF	ARFIMA-MF
1	0.983	0.985	0.981	0.981	0.984	0.999
5	0.975	0.955	0.932	0.993	0.983	0.997
10	0.940	0.921	0.903	0.993	0.985	0.998
20	0.920	0.908	0.898	0.988	0.982	0.998
30	0.925	0.929	0.926	0.980	0.976	0.999
40	0.908	0.917	0.911	0.973	0.968	0.998
50	0.855	0.869	0.857	0.966	0.959	0.998
60	0.751	0.764	0.762	0.960	0.956	0.999
70	0.461	0.502	0.501	0.951	0.946	0.999
80	0.278	0.317	0.323	0.943	0.939	0.999
90	0.143	0.190	0.201	0.939	0.933	0.998
100	0.060	0.109	0.118	0.930	0.926	0.999
	MAE					
lead	GARCH-FIGARCH	GARCH - ARFIMA	GARCH-MF	FIGARCH-ARFIMA	FIGARCH -MF	ARFIMA-MF
1	0.991	0.983	0.984	0.995	0.993	0.997
5	0.977	0.961	0.962	0.994	0.990	0.994
10	0.955	0.925	0.933	0.987	0.988	0.991
20	0.889	0.823	0.865	0.975	0.983	0.985
30	0.758	0.646	0.732	0.953	0.975	0.976
40	0.737	0.640	0.722	0.945	0.973	0.980
50	0.696	0.560	0.647	0.925	0.962	0.976
60	0.711	0.545	0.643	0.913	0.959	0.976
70	0.655	0.504	0.589	0.912	0.959	0.977
80	0.594	0.513	0.574	0.899	0.950	0.978
90	0.490	0.494	0.519	0.889	0.939	0.983
100	0.357	0.395	0.399	0.869	0.930	0.979

Note: At a significance level of 95 percent, the null hypothesis of no correlation in the performance of different methods would have to be rejected for absolute entries beyond 0.197 (= 1.96 √(n-1)) with n = 100 in our case).

Table 15: Rank Correlations Across Assets: Forecasts of Volume, Random Sample

	MSE			MAE		
lead	ARMA-ARFIMA	ARMA-MF	ARFIMA-MF	ARMA-ARFIMA	ARMA-MF	ARFIMA-MF
1	0.926	0.875	0.989	0.901	0.898	0.997
5	0.640	0.647	0.996	0.481	0.556	0.985
10	0.761	0.776	0.994	0.600	0.669	0.986
20	0.493	0.517	0.993	0.609	0.644	0.974
30	0.347	0.355	0.994	0.517	0.509	0.966
40	0.160	0.163	0.992	0.389	0.330	0.958
50	0.084	0.089	0.991	0.301	0.185	0.950
60	0.021	0.024	0.991	0.270	0.126	0.938
70	-0.039	-0.040	0.988	0.243	0.089	0.930
80	-0.070	-0.076	0.986	0.238	0.081	0.922
90	-0.085	-0.084	0.985	0.229	0.061	0.916
100	-0.090	-0.096	0.980	0.218	0.036	0.906

Note: At a significance level of 95 percent, the null hypothesis of no correlation in the performance of different methods would have to be rejected for absolute entries beyond 0.197 (= 1.96 √(n-1)) with n = 100 in our case).

5. Conclusion

This paper has examined the potential of time series models with long memory (FIGARCH, ARFIMA, multi-fractal) to improve upon the forecasts derived from short-memory models (GARCH for volatility, ARMA for volume). In order to get a broad picture, we have used a large data-base applying the competing models to long forecasting horizons for a long out-of-sample period. A number of interesting results emerged form this exercise: first, as concerns volatility, our selection of long-memory models performs better in most cases than the naive sample variance and GARCH forecasts.

However, this potential improvement against short-memory models is overshadowed by occasional dramatic failures particularly by the FIGARCH model and to lesser extent by ARFIMA. Interestingly, the newly proposed multi-fractal approach seems not to suffer at all from this problem. Remarkably, results are better throughout for the MSE than the MAE criterion (some trial runs with other data suggest that this is not a particularity of the Japanese market). Time series methods, thus, seem to be better suited for forecasting large realizations of volatility rather than small or medium ones.

Second, as concerns volume, we find a higher degree of forecastability than with volatility (both under the MSE and MAE criterion) and again a dominance of long-memory models (ARFIMA and MF). As with volatility, MF also provides much safer forecasts which never rise above the benchmark of unity under the relative MSE criterion.

Third, our observation of different degrees of the variability of performance of different methods motivated an analysis of the forecasting quality of pooled estimates (i.e. mean values of estimated parameters over the one hundred selected stocks). Astonishingly, nothing was lost by discarding stock-specific estimates, but results improved under practically all perspectives. In particular, the formerly more 'dangerous' methods with some extremely poorly performing cases now also became as safe as the multi-fractal model. Using pooled estimates, we also see an even more clear-cut difference between all long-memory models and their short-memory counterparts: as can be seen in *Figs. 4* and *5*, pooled GARCH and ARMA quickly converge to the behaviour of naïve forecasts for increasing forecasting horizon yielding uniform relative MSE and relative MAE equal to unity from horizons of about forty days. In contrast, the long-memory models all have a uniformly better performance at least with respect to the MSE criterion. In our view, this finding speaks in favour of 'true' long-term dependence being present in the data. It is also remarkable, that rank correlations over markets in the pooled estimation cases are close to unity for the long-memory models showing that they all extract similar information. Overall, these results suggest the following interpretation: volatility and volume are characterized by processes which have strong persistency. This persistent component can be captured to some degree by different time series models which have built-in long correlations. The improvement form pooled estimates indicates that persistency is similar across stocks so that one gets a better assessment of the dependence structure by increasing the data size via merging all stocks rather than by the more common fine-tuning of individual estimates. Work in progress, in fact, confirms this view: in preliminary experiments we applied our average estimates from the Japanese market to data from other countries and again found a better performance than with individually estimates parameters. A more systematic exploration of these findings is left for future research.

References:

Baillie, R.T., T. Bollerslev and H.O. Mikkelsen (1996) Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics* 74, 3 – 30.

Basak, G., N. Chan and W. Palma (2001) The Approximation of Long-Memory Processes by an ARMA Model, *Journal of Forecasting* 20, 367-389.

Beran, J. (1994) *Statistics for Long-Memory Processes*. New York: Chapman & Hall.

Bollerslev, T. and D. Jubinski (1999) Equity Trading Volume and Volatility: Latent Information Arrivals and Common Long-Run Dependencies, *Journal of Business & Economic Statistics* 17, 9-21.

Brailsford, T. and R. Fall (1996) An Evaluation of Volatility Forecasting Techniques, *Journal of Banking and Finance* 20, 419-438

Brockwell, P.J. and R.A. Davis (1991) *Time Series: Theory and Methods*. 2nd ed., New York, Springer.

Brodsky, J. and C. Hurvich (1999) Multi-Step Forecasting for Long-Memory Processes, *Journal of Forecasting* 18, 59-75.

Calvet, L. and A. Fisher (2001) Forecasting Multifractal Volatility, *Journal of Econometrics* 105, 27 – 58.

Calvet, L. and A. Fisher (2002) Multi-Fractality in Asset Returns: Theory and Evidence, *Review of Economics & Statistics* 84, 381 – 406.

Calvet, L. and A. Fisher (2003) *Regime-Switching and the Estimation of Multifractal Processes*. Mimeo: Harvard University.

Chong, Y. and D. Hendry, Econometric Evaluation of Linear Macroeconomic Models, *Review of Economic Studies* 53, 1986, 671 – 90.

Chung, C.-F. (2002) *Estimating the Fractionally Integrated GARCH Model*. Mimeo: Academia Sinica.

Crato, N. and B. Ray (1996) Model Selection and Forecasting for Long-Range Dependent Processes, *Journal of Forecasting* 15, 107-125.

Dimson, E. and P. Marsh (1990) Volatility Forecasting without Data-Snooping, *Journal of Banking and Finance* 14, 399-421.

Ding, Z., C.W.J. Granger and R.F. Engle (1993) A Long Memory Property of Stock Market Returns and a New Model, *Journal of Empirical Finance* 1, 83 - 106.

Figlewski, S. (1997) Forecasting Volatility, *Financial Markets, Institutions & Investment* 6, 1-88.

Fisher, A., L. Calvet and B. Mandelbrot (1997) *Multifractality of Deutschemark/US Dollar Exchange Rates*. Mimeo: Cowles Foundation for Research in Economics.

Fong, W. (1997) Volatility Persistence and Switching ARCH in Japanese Stock Returns, *Financial Engineering and the Japanese Markets* 4, 37-57.

Granger, C. and R. Joyeux (1980) An Introduction to Long-Memory Time Series Models and Fractional Differencing, *Journal of Time Series Analysis* 1, 1980, 15 – 39.

Kaastra, I. And M. Boyd (1995) Forecasting Futures Trading Volume Using Neural Networks, *Journal of Futures Markets* 15, 1995, 953 – 970.

Klaassen, F. (2002) Improving GARCH Volatility Forecasts with Regime-Switching GARCH, *Empirical Economics* 27, 363-394.

Lux, T. (2003) The Multi-Fractal Model of Asset Returns: Its Estimation via GMM and its Use for Volatility Forecasting. *Working paper*, University of Kiel.

Man, K. (2003) Long Memory Time Series and Short Term Forecasts, *International Journal of Forecasting* 19, 477-491.

Mandelbrot, B. (1974) Intermittent Turbulence in Self Similar Cascades: Divergence of High Moments and Dimension of the Carrier, *Journal of Fluid Mechanics* 62, 331 – 358.

Mandelbrot, B. (1999) A Multifractal Walk down Wall Street, *Scientific American*, February, 50 - 53.

Mandelbrot, B., A. Fisher and L. Calvet (1997) *A Multifractal Model of Asset Returns*. mimeo: Cowles Foundation for Research in Economics.

Poon, S.-H. and C. Granger (2003) Forecasting Volatility in Financial Markets: A Review, *Journal of Economic Literature*, in press.

Ray, B., S. Chen and J. Jarrett (1997) Identifying Permanent and Temporary Components in Daily and Monthly Japanese Stock Prices, *Financial Engineering and the Japanese Markets* 4, 233-256.

Tse, Y. (1991) Stock Returns Volatility in the Tokyo Stock Exchange, *Japan and the World Economy* 3, 285-298.

Vilasuso, J. (2002) Forecasting Exchange Rate Volatility, *Economics Letters* 76, 59 – 64.

West, K. and D. Cho (1995) The Predictive Ability of Several Models of Exchange Rate Volatility, *Journal of Econometrics* 69, 367-391.

Zumbach, G. (2004) Volatility Processes and Volatility Forecast with Long Memory, *Quantitative Finance* 4, 70 – 86.

Appendix: Stocks used in the analysis

The 'random sample' consisted of the following stocks which were randomly chosen from our data base (in parentheses: stock identification number):

Nippon Suisan Kaisha (1332), Hoko Fishing (1351), Fudo Construction (1813),Tekken Corp. (1815), Nakano Corp. (1827), Toda Corp. (1860), Penta-Ocean Construction (1893), Obayashi Road Corp. (1896),Daiwa House Industry (1925), Nippon Koei (1954), Morinaga (2201), Nippon Meat Packers (2282), Itoham Foods (2284), Nichirei Corp. (2871), Kawashima Textile Manufacturers (3009), Nisshinbo Industries (3105), Ashimori Industries (3526), Showa Denko K.K. (4004), Nippon Carbide Industries (4064), Sakai Chemical Industries (4078, until 09/97), Mitsui Chemicals (4183), JSR Corp. (4185), Nippon Kayaku (4272), Sankyo (4501), Yamanouchi Pharmaceutical (4503), Daiichi Pharmaceutical (4505), Kansai Paint (4613), Tohpe Corp. (4614), Chugoku Marine Paints (4617), Toyo Ink Mfg. (4634),Takasago International Corp. (4914), Toyo Tire & Rubber (5105), Osaka Cement (5235, until 12/93), Nippon Hume Corporation (5262),Nippon Yakin Kogyo (5480), Nippon Denko (5563), Suzuki Metal Industries (5657),Nihon Seiko (5729), Sakurada (5917), Amada Machines (6107), Koike Sanso Kogyo (6137), Kioritz Corp. (6313), Meiji Machines (6334), Sintokogio (6339), Sumitomo Precision Products (6355), Nippon Gear (6356), Sakai Heavy Industries (6358), Toyo Kanetsu K.K. (6369), Tsubakimoto Chain Co. (6371), Fuso Lexel (6386), Sanjo Machine Works (6437), Riken Corp. (6462), Koyo Seiko (6473), Yashakaw Electric Corp. (6506), Makita Corp .(6586), Nishishiba Electric (6591), Kawaden (6648, until 09/2000), The Nippon Signal (6741), Nec Tokin Corp. (6759), Kawasaki Heavy Industries (7012), Shin Maywa Industries (7224), Tokyo Radiator Mfg. (7235), Daihatsu Motor (7262), Oval Corp. (7727), Riken Keiki (7734), Chinon Industries (7738), Pentax Corp. (7750), Canon (7751), Dantani Corp. (7910), Yamaha Corp. (7951), Takara Standard (7981), Daiwa Seiko (7990), Kanematsu Corp. (8020), Tohto Suisan (8038), Tsukiji Uoichiba (8039), Seiko Corp. (8050), Shoko (8090), Inabata (8098), GSI Creos Corp. (8101), Sinanen (8132), Matsuzakaya (8235), Maruzen (8236), Bank of Yokohama (8332), Gunma Bank (8334), Musashino Bank (8336), Hyakugo Bank (8368), Kiyo Bank (8370), Iyo Bank (8385), Oita Bank (8392), Sumitomo Insurance (8753, until 09/2001), Nipponkoa Insurance (8754), Sompo Japan Insurance Inc. (8755), Nissan Fire & Marine Ins. (8756), Nissay Dowa General Insurance (8759), Nichido Fire & Marine Ins. (8760), Taiheiyo Kaiun (9123), KDD (9431, until 09/2000), Chugoku Electric Power (9504), Toho Gas (9533), Shochiku (9601).

The 100 stocks with the highest average trading volume are (stock identification numbers in parentheses):

Teikoku Oil (1601), Taisei Corp. (1801), Obayashi Corp. (1802), AC Real Estate Corp. (1806), Kajima Corp. (1812), Kumagai Gumi (1861), Aoki Corp. (1886), Daiwa House Industry (1925), Kirin Brewery (2503), Toyobo (3101), Unitika (3103), Teijin (3401), Toray Industries (3402), Mitsubishi Rayon (3404), Asahi Kasei Corp. (3407), Sangoku Pulp (3702, until 07/92), Oji Paper (3861), Mitsui Toatsu Chemical (4001, until 12/96), Showa Denko K.K. (4004), Sumitomo Chemical (4005), Mitsubishi Chemical Corp. (4010), Ishihara Sangyo Kaisha (4028), Tosoh Corp. (4042), Denki Kagaku Kogyo Kabushiki Kai (4061), Ube Industries (4208), Takeda Chemical Industries (4502), Dainippon Ink and Chemicals (4631), Fuji Photo Film (4901), Nippon Oil Corp. (5001), Mitsubishi Oil (5004, until 03/99), Cosmo Oil (5007), Nippon Sheet Glass (5202), Taiheiyo Cement Corp. (5233), Nippon Steel Corp. (5401), Kawasaki Steel (5403), NKK Corp. (5404), Sumitomo Metal Industries (5405), Kobe Steel (5406), Nisshin Steel (5407), The Japan Steel Works (5631), Nippon Light Metal Company (5701), Mitsui Mining and Smelting (5706), Mitsubishi Materials Corp. (5711), Nippon Mining (5712, until 07/91), Sumitomo Metal Mining (5713), Dowa Mining (5714), The Furukawa Electric (5801), Sumitomo Electric Industries (5802), Fujikura (5803), Komatsu (6301), Sumitomo Heavy Industries (6302), Kubota Corp. (6326), Hitachi (6501), Toshiba Corp. (6502), Mitsubishi Electric Corp. (6503), Fuji Electric (6504), Nec Corp. (6701), Fujitsu (6702), Oki Electric Industry (6703), Matsushita Electric Industrial (6752), Sharp Corp. (6753), Sony Corp. (6758), Sanyo Electric (6764), Mitsui Engineering & Shipbuilding (7003), Hitachi Zosen Corp. (7004), Mitsubishi Heavy Industries (7011), Kawasaki Heavy Industries (7012), Ishikawajima-Harima Heavy Indust. (7013), Nissan Motor (7201), Isuzu Motors (7202), Toyota Motor Corp. (7203), Mazda Motor Corp. (7261), Honda Motor (7267), Fuji Heavy Industries (7270), Canon (7751), Ricoh Company (7752), Itochu Corp. (8001), Marubeni Corp. (8002), Mitsui (8031), Mitsubishi Corp. (8058), Nissho Iwai Corp. (8063), Sakura Bank (8314, until 03/2001), Bank of Tokyo-Mitsubishi (8315, until 03/2001), Sumitomo Bank (8318), Asahi Bank (8322), Daiwa Securities Group Inc. (8601), Yamaichi (8602, until 11/97), Nikko Cordial Corp. (8603), Nomura Holdings (8604), Tokio Marine & Fire Ins. (8751), Mitsui Fudosan (8801), Mitsubishi Estate (8802), Tobu Railway (9001), Tokyu Corp. (9005), Keisei Electric Railway (9009), Mitsui O.S.K. Lines (9104), Kawasaki Kisen Kaisha (9107), The Tokyo Electric Power (9501), Tokyo Gas (9531), Osaka Gas (9532)