TI 2004-067/4
Tinbergen Institute Discussion Paper

Modeling and Forecasting S&P 500 Volatility: Long Memory, Structural Breaks and Nonlinearity

Martin Martens
Dick van Dijk
Michiel de Pooter

Faculty of Economics, Erasmus Universiteit Rotterdam, and Tinbergen Institute.

Tinbergen Institute
The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam, and Vrije Universiteit Amsterdam.

Tinbergen Institute Amsterdam
Roetersstraat 31
1018 WB Amsterdam
The Netherlands
Tel.: +31(0)20 551 3500
Fax: +31(0)20 551 3555

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Modeling and Forecasting S&P 500 Volatility: Long Memory, Structural Breaks and Nonlinearity*

Martin Martens[†]
Econometric Institute
Erasmus University Rotterdam

Dick van Dijk[‡]
Econometric Institute
Erasmus University Rotterdam

Michiel de Pooter[§]
Tinbergen Institute
Erasmus University Rotterdam

June 5, 2004

Abstract

The sum of squared intraday returns provides an unbiased and almost error-free measure of ex-post volatility. In this paper we develop a nonlinear Autoregressive Fractionally Integrated Moving Average (ARFIMA) model for realized volatility, which accommodates level shifts, day-of-the-week effects, leverage effects and volatility level effects. Applying the model to realized volatilities of the S&P 500 stock index and three exchange rates produces forecasts that clearly improve upon the ones obtained from a linear ARFIMA model and from conventional time-series models based on daily returns, treating volatility as a latent variable.

Key words: Realized volatility, high-frequency data, long memory, day-of-the-week effect, leverage effect, volatility forecasting, smooth transition autoregression.

JEL Classification Code: C22, C53, G15

---

*We would like to thank seminar participants at the "New Frontiers in Financial Volatility Modelling" conference in Florence (May 2003) and the CIRNAO-CIREQ conference on "Realized Volatility" in Montreal (November 2003) for providing useful comments.

†Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, NL-3000 DR Rotterdam, The Netherlands, e-mail: `mmartens@few.eur.nl`

‡Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, NL-3000 DR Rotterdam, The Netherlands, e-mail: `djvandijk@few.eur.nl` (corresponding author)

§Tinbergen Institute, Erasmus University Rotterdam, P.O. Box 1738, NL-3000 DR Rotterdam, The Netherlands, e-mail: `depooter@few.eur.nl`

1 Introduction

Accurately measuring and forecasting financial volatility is of crucial importance for asset and derivative pricing, asset allocation and risk management. Merton (1980) already noted that the variance over a fixed period can be estimated arbitrarily accurately by the sum of squared intra-period realizations, provided the data are available at a sufficiently high sampling frequency. With transaction prices becoming more widely available, Andersen and Bollerslev (1998) kick-started a flurry of research on the use of high-frequency data for measuring and forecasting volatility. Andersen and Bollerslev (1998) showed that ex-post daily exchange rate volatility is best measured by aggregating 288 squared five-minute returns. The five-minute frequency is a trade-off between accuracy, which is theoretically optimized using the highest possible frequency, and noise due to, for example, the bid-ask bounce.[1] Ignoring the small remaining measurement error the volatility essentially becomes "observable" ex-post.[2] As such, volatility can be modeled directly, rather than being treated as a latent variable as is the case in GARCH and stochastic volatility models. The main drawback of such models is the need to make specific assumptions regarding the distribution of shocks and the properties of the latent volatility factor. The sum of intraday squared returns is also a much more accurate measure of daily realized volatility than the popular daily squared return.[3]

Several recent studies document the properties of realized volatilities constructed from high-frequency data for different financial assets, including exchange rates (Andersen, Bollerslev, Diebold and Labys, 2001), stock indexes and corresponding futures (Ebens, 1999; Areal and Taylor, 2002; Martens, 2002; Thomakos and Wang, 2003) and individual stocks (Andersen, Bollerslev, Diebold and Ebens ,2001). One of the most important stylized facts to come out of these studies is that realized volatilities are fractionally integrated of order d, where d typically is around 0.4. This property is used for modeling and forecasting volatilities at daily or longer horizons for both exchange rates (Andersen, Bollerslev, Diebold and Labys, 2003; Li, 2002; Pong, Shackleton, Taylor and Xu, 2004) and stock indexes (Ebens, 1999; Hol, Jungbacker and Koopman, 2004; Martens and Zein, 2004). These studies use

[1]See Zhang, Mykland and Aït-Sahalia (2003), Aït-Sahalia, Mykland and Zhang (2004) and Bandi and Russell (2003, 2004) for recent discussions involved in the choice of optimal sampling frequency.

[2]See Andersen, Bollerslev and Diebold (2002) and Barndorff-Nielsen and Shephard (2002a,b, 2003, 2004a,b) for formal discussions of the theoretical properties of realized volatility and the related concept of power variation.

[3]More accurate in the sense that it has smaller variance. Also, of relevance to this study, estimates of the degree of fractional integration are unbiased for daily volatility based on intraday returns, whereas they are severely downward biased when estimated from daily squared returns, see Bollerslev and Wright (2000).

Autoregressive Fractionally Integrated Moving Average (ARFIMA) models to capture both the long memory characteristic and any remaining short-term dynamics. The resulting forecasts generally outperform those obtained from ARCH models (including GARCH, EGARCH and FIGARCH), Riskmetrics' historical volatility with exponentially declining weights, and stochastic volatility models, and they can compete with implied volatility forecasts obtained from options. The latter is noteworthy as the current literature (e.g. Jorion (1995) and Christensen and Prabhala (1998)) suggests that implied volatility forecasts are superior to forecasts obtained from ARCH models, to the extent that ARCH forecasts do not contain any information not already subsumed by implied volatility forecasts.

Except for Ebens (1999) all the aforementioned studies use linear models, which ignore several empirically important aspects of financial volatility. First, linear models do not allow for the so-called leverage effect documented by, among others, Black (1976), Pagan and Schwert (1990) and Engle and Ng (1993). These studies show an asymmetric relation between news (as measured by lagged unexpected returns) and volatility, in that a negative return tends to increase subsequent volatility by more than would a positive return of the same magnitude. Second, Longin (1997) reports that shocks may be less persistent in periods of high volatility than in case volatility is low. Third, occasional structural breaks can spuriously suggest the presence of long memory, as shown by Diebold and Inoue (2001), among others. As financial volatility has been found to experience irregular level shifts (see Lamoureux and Lastrapes, 1990, and Andreou and Ghysels, 2002), it seems important to consider this characteristic when modeling realized volatility. Fourth, Baillie and Bollerslev (1989) and Harvey and Huang (1991), among others, find that average volatility is not constant across the different days of the week but rather displays a pronounced U-shaped pattern with volatility being lowest on Wednesdays.

In this paper, we propose a nonlinear model for realized volatilities that simultaneously captures long memory, leverage effects, volatility persistence that depends on the size of the shock, structural breaks and day-of-the-week effects. To the best of our knowledge, we are the first to develop such a comprehensive nonlinear model. The small number of previous studies that have considered nonlinearities in realized volatilities all are limited in one way or another. Ebens (1999), Oomen (2002) and Giot and Laurent (2004) incorporate leverage effects in a long memory model for various stock indexes. Only Giot and Laurent (2004) consider out-of-sample forecasting, but only at the one-day horizon. Maheu and McCurdy (2002) use a regime-switching model for the DM/$ exchange rate, but do not consider any other nonlinearities or long memory and only forecast one-day-ahead.

Our model is estimated and used to produce volatility forecasts at various horizons for S&P 500 index-futures and three exchange rates, the DM/$, ¥/$ and ¥/DM. The S&P results first of all show that level shifts in S&P 500 volatility do not account for the long memory feature. The fractional integration parameter does decline when explicitly modeling the structural break, but remains significantly different from zero. Second, the day-of-the-week dummies show that volatility is on average lower on Mondays and Tuesdays and higher on Fridays. This is an interesting contrast with the U-shaped pattern found in daily squared returns, which also attribute a higher volatility to Mondays and Tuesdays. Third, we find convincing evidence for the presence of a leverage effect in S&P volatility, in that negative lagged returns significantly increase volatility whereas positive returns do not affect volatility at all. Incorporating these nonlinear features is important for out-of-sample forecasting as well. We find that 1-day-ahead volatility forecasts from the best nonlinear model improve upon those from a linear ARFIMA model on all evaluation criteria considered. For example, the R^2 from a regression of realized volatility on the volatility forecast increases from 42.1% to 46.1%. For the exchange rates the leverage effect is less important, as expected, but incorporating nonlinearities still improves the in-sample fit and out-of-sample forecast performance. For the ¥/$, for example, the R^2 from a regression of realized volatility on the 1-day-ahead volatility forecast increases substantially from 36.6% to 55.4%.

In the sequel, we first focus fully on the S&P 500, partly for expositional convenience, and partly because the exchange rate data are the same as in Andersen, Bollerslev, Diebold and Labys (2001) and analysed thoroughly there.[4] First, we discuss the S&P 500 data in Section 2. The nonlinear long-memory model is developed in Section 3. We discuss estimation and forecasting results for the S&P 500 in Sections 4 and 5, respectively. Section 6 then summarizes the most important findings for the exchange rates. Finally, Section 7 concludes.

2 Data

We construct our measure of daily realized volatility for the S&P 500 index using high-frequency futures data. S&P 500 index futures trade on the Chicago Mercantile Exchange (CME) on the trading floor from 8:30AM to 3:15PM (Eastern Standard Time minus 1 hour, EST-1). Since January 3, 1994, these contracts also trade overnight on GLOBEX, the electronic trading system of the CME, from 3:30PM to 8:00AM (8:15AM from February 26, 1996, onwards). As a result, S&P 500 futures trade almost round the clock, providing a similar opportunity to construct realized

[4]We would like to thank Torben Andersen for providing the daily exchange rate returns and corresponding realized volatility series.

volatilities as for the 24-hour FX market. Martens (2002) tested various measures of S&P 500 realized volatility, finding that the sum of squared 30-minute intranight and 5-minute intraday returns is a more accurate measure of volatility than using only the intraday returns, or the sum of squared intraday returns and the squared close-to-open return, showing that it is useful to incorporate overnight trading prices. Hence, we will use the following measure of daily "realized volatility",

$$s_t^2 = \sum_{j=1}^{n_N} \left(r_{t,j}^N\right)^2 + \sum_{j=1}^{n_D} \left(r_{t,j}^D\right)^2, \tag{1}$$

where $r_{t,j}^N$ is the intranight (30-minute) return on day t in intranight period j ($j = 1, \ldots, n_N = 33$), and $r_{t,j}^D$ is the intraday (five-minute) return on day t for intraday period j ($j = 1, \ldots, n_D = 91$).

Figure 1 shows time series plots for the daily S&P 500 realized volatility, realized standard deviation, and the log realized volatility for the sample period from January 3, 1994, until December 29, 2000 (1767 daily observations). Table 1 contains descriptive statistics of these realized volatility measures, as well as for daily returns $r_t = \sum_{j=1}^{n_N} r_{t,j}^N + \sum_{j=1}^{n_D} r_{t,j}^D$, for squared and absolute daily returns, and for daily returns standardized with the realized standard deviation r_t/s_t. A number of interesting features emerge from this table, which closely correspond with the distributional characteristics for realized exchange rate volatility documented in Andersen, Bollerslev, Diebold and Labys (2001). First, comparing the daily squared returns with the realized variance shows that both have almost the same mean (1.217% and 1.194%, respectively). We would expect this to be the case, as both are unbiased measures of the true volatility. However, the standard deviation of the realized variance is at 1.770 much smaller than the standard deviation of the squared returns, which equals 3.242. It is precisely this characteristic that shows that realized variance is a much less noisy estimate of true volatility than the daily squared return. Second, the realized variance and realized standard deviation are heavily skewed and exhibit excess kurtosis. By contrast, the logarithm of realized volatility, $\log(s_t^2)$, is much more symmetrically distributed and has much lower kurtosis. This is corroborated by the kernel density estimates shown in Figure 2, from which it is seen that log realized volatility is approximately normally distributed. It is for this reason that we will consider time series models for the log realized volatility. Third, the daily S&P 500 returns are skewed and leptokurtic; see also panel (a) of Figure 3, which shows that the returns distribution is peaked and fat-tailed. By contrast, the standardized returns r_t/s_t exhibit much less skewness and excess kurtosis and in fact are very close to being normally distributed, see panel (b) of Figure 3.

- insert Table 1 and Figures 1-3 about here -

Fourth, Figure 4 shows sample autocorrelation functions for daily squared returns and absolute returns, and for daily realized variance, realized standard deviation and log realized variance. The autocorrelations for the realized volatility measures exhibit a slow hyperbolic decay, indicative for the presence of long memory. Note that the persistence in the autocorrelation functions for the realized volatility measures is much stronger than for the daily squared and absolute returns.

- insert Figure 4 about here -

Fifth, returning to Figure 1, realized volatility appears to be higher on average at the end of the sample period than during the first few years. It is difficult to pin down when exactly this level shift occurred, and it appears that it is most adequately characterized as a gradual increase of volatility during 1996-1997. An alternative possibility is that multiple structural breaks have occurred, as suggested by Andreou and Ghysels (2002).

The scatter plot of $\log(s_t^2)$ against r_{t-1} in Figure 5 reveals a rather pronounced relationship between current volatility and lagged returns. To examine the possible presence of a leverage effect, we estimate the "news impact curve" (Engle and Ng, 1993)

$$\log(s_t^2) = \beta_0 + \beta_1|r_{t-1}| + \beta_2\mathtt{I}[r_{t-1} < 0] + \beta_3|r_{t-1}|\mathtt{I}[r_{t-1} < 0], \tag{2}$$

where $\mathtt{I}[A]$ is an indicator function for the event A, being equal to 1 if A occurs, and 0 otherwise. The fit from this regression is included in Figure 5 as well, along with the fit from a symmetric version of this news impact curve, obtained by setting $\beta_2 = \beta_3 = 0$ in (2). It is clearly seen that the impact of negative lagged returns is larger than the effect of positive returns of equal magnitude. Also, the parametric form in (2) appears to be quite reasonable, as can be seen by comparing the fit from this regression with a nonparametric regression of log realized volatility on the lagged return, also shown in Figure 5.

- insert Figure 5 about here -

Finally, Table 2 shows the overall mean and the mean on different days-of-the-week for all return and volatility measures. The common finding based on daily returns that Mondays and Fridays exhibit higher volatility than other days is confirmed by the S&P data. Interestingly, this pattern is quite different for the realized variance. Thursdays and Fridays exhibit the highest volatility, and Mondays no longer have an above average volatility. For the log realized variance this is even more pronounced: its mean is lowest on Mondays and monotonically increases during the week. This is most likely due to daily observations using the Friday close to

Monday close return, whereas with transaction data it is the Friday close to Sunday evening start of overnight futures trading, which then gradually adjusts to information over the weekend. A large (Friday to Monday) return will always result in a large daily squared return, whereas transaction data distinguish a gradual price change from a truly volatile day with large swings.

- insert Table 2 about here -

3 Nonlinear Long Memory Models

Following previous studies, we employ Autoregressive Fractionally Integrated Moving Average (ARFIMA) models to describe the dynamic properties of logarithmic realized volatility $y_t = \log(s_t^2)$,

$$\phi(L)(1-L)^d(y_t - \mu_t) = \varepsilon_t, \tag{3}$$

where the order of integration d is allowed to take non-integer values, $\phi(L) = 1 - \phi_1 L - \ldots - \phi_p L^p$ is a p-th order lag polynomial assumed to have all roots outside the unit circle and ε_t is a white noise process. It is common practice to set the mean μ_t equal to a constant, i.e. $\mu_t = \alpha_0$. However, to capture the salient features of the S&P realized volatility discussed in the previous section, we extend the model to allow for gradual level shifts, day-of-the-week effects, and nonlinear effects of lagged returns by setting

$$\begin{aligned}\mu_t = \alpha_0 + P(t) + \beta_1|r_{t-1}| + \beta_2\mathtt{I}[r_{t-1} < 0] + \beta_3|r_{t-1}|\mathtt{I}[r_{t-1} < 0]& \\ + \delta_1 D^*_{1,t} + \delta_2 D^*_{2,t} + \delta_4 D^*_{4,t} + \delta_5 D^*_{5,t}&\end{aligned} \tag{4}$$

where $D^*_{s,t} \equiv D_{s,t} - D_{3,t}$ $D_{s,t}$, $s = 1, 2, 4, 5$ are "centered" daily dummy variables, with $D_{s,t} = 1$ when time t corresponds with day s (1=Monday, 2=Tuesday, etc.) and $D_{s,t} = 0$ otherwise, and where $P(t)$ is a fifth-order polynomial in t to capture gradual level shifts in volatility,

$$P(t) = \alpha_1 t + \alpha_2 t^2 + \alpha_3 t^3 + \alpha_4 t^4 + \alpha_5 t^5, \tag{5}$$

We also considered an alternative approach by replacing $P(t)$ by $\alpha_1 G(t; \gamma, \tau)$ where $G(t; \gamma, \tau)$ is the logistic function that changes monotonically from 0 to 1 as t increases[5]. Although the latter approach yields similar results, it can be too restrictive since it only allows for a single structural change in the level of y_t. The above ARFI

[5]For further details about the logistic function see the discussion below.

model with Structural Change (SC), Dummies (D), lagged Return (R) and Leverage effects (L) will be denoted ARFI-SCDRL.

We also estimate an alternative model, where (cf. Ebens, 1999) terms involving the lagged returns are not included in the conditional mean μ_t, but as "exogenous regressors" (X), leading to the model (denoted ARFI-SCDXRL)

$$\phi(L)(1-L)^d(y_t-\mu_t)=\tilde{\beta}_1|r_{t-1}|+\tilde{\beta}_2\mathtt{I}[r_{t-1}<0]+\tilde{\beta}_3|r_{t-1}|\mathtt{I}[r_{t-1}<0]+\varepsilon_t, \qquad (6)$$

where now

$$\mu_t=\alpha_0+P(t)+\delta_1 D^*_{1,t}+\delta_2 D^*_{2,t}+\delta_4 D^*_{4,t}+\delta_5 D^*_{5,t}. \qquad (7)$$

Finally, we examine whether the persistence of shocks depends on the level of volatility (cf. Longin, 1997) by allowing for regime switching behavior in the short-run dynamics of realized volatility. Specifically, we estimate a long-memory smooth transition autoregressive model (denoted STARFI-SCDXRL),

$$\begin{aligned}(\phi_1(L)(1-G(s_t;\gamma,c))+\phi_2(L)G(s_t;\gamma,c))(1-L)^d(y_t-\mu_t)&= \qquad (8)\\ \tilde{\beta}_1|r_{t-1}|+\tilde{\beta}_2\mathtt{I}[r_{t-1}<0]+\tilde{\beta}_3|r_{t-1}|\mathtt{I}[r_{t-1}<0]+\varepsilon_t&\end{aligned}$$

where $\phi_1(L)=1-\phi_{1,1}L-\ldots-\phi_{1,p}L^p$ and $\phi_2(L)=1-\phi_{2,1}L-\ldots-\phi_{2,p}L^p$ are p-th order lag polynomials. $G(s_t;\gamma,c)$ is the logistic function

$$G(s_t;\gamma,c)=\frac{1}{1+\exp\{-\gamma(s_t-c)/\sigma_{s_t}\}}\text{ with }\gamma>0 \qquad (9)$$

where s_t is the transition variable and σ_{s_t} is the standard deviation of s_t. As s_t increases, the logistic function changes monotonically from 0 to 1, with the change being symmetric around the location parameter c, as $G(c-z;\gamma,c)=1-G(c+z;\gamma,c)$ for all z. The slope parameter γ determines the smoothness of the change in the value of the logistic function. As $\gamma\to\infty$, $G(s_t;\gamma,c)$ approaches the indicator function $\mathtt{I}[s_t>c]$ and, consequently, the change of $G(s_t;\gamma,c)$ from 0 to 1 becomes instantaneous at $s_t=c$. When $\gamma\to 0$, $G(s_t;\gamma,c)\to 0.5$ for all values of s_t. In that case there is no regime-switching behavior in the autoregressive parameters and (8) then reduces to (6). Given our objective to allow for different persistence in periods of high and low volatility, we use the one-period lagged volatility in deviation from the estimated polynomial as the transition variable in (8), i.e. $s_t=y_{t-1}-P(t-1)$ with $P(t)$ given by (5).

To gauge the relative importance of the different nonlinear features of realized volatility, we also estimate several restricted versions of the full models. In particular, we consider (i) a model without the leverage effect but including the lagged absolute return ($\beta_2=\beta_3=0$ in (4) or $\tilde{\beta}_2=\tilde{\beta}_3=0$ in (6); ARFI-SCD(X)R), (ii) a model without any effect of lagged returns at all ($\beta_1=\beta_2=\beta_3=0$ in (4) or $\tilde{\beta}_1=\tilde{\beta}_2=$

$\tilde{\beta}_3 = 0$ in (6); ARFI-SCD), and (iii) a model without any effect of lagged returns and without day-of-the-week effects (imposing in addition $\delta_1 = \delta_2 = \delta_4 = \delta_5 = 0$ in (4) or (7); ARFI-SC). Finally, we also estimate all models without allowing for a structural change, by imposing $\alpha_i = 0$ for $i = 1, \ldots, 5$ in (5).

For estimation of the parameters in the ARFI models we use Beran's (1995) approximate maximum likelihood (AML) estimator for invertible and possibly non-stationary ARFIMA models (i.e. for $d > -0.5$), which amounts to minimizing the sum of squared residuals

$$Q_n(\theta) = \sum_{t=1}^{T} e_t^2(\theta), \tag{10}$$

where $\theta = (d, \gamma, \tau, \alpha, \beta, \delta, \phi)$, T is the sample size and the residuals $e_t(\theta)$ are computed as

$$e_t(\theta) = (y_t - \mu_t) - \sum_{j=1}^{t+p-1} \pi_j (y_{t-j} - \mu_{t-j}) \tag{11}$$

where the π_j are the autoregressive coefficients in the infinite order AR representation of the ARFI models

$$\pi(L)(y_t - \mu_t) = \varepsilon_t, \tag{12}$$

that is $\pi(L) = 1 - \pi_1 L - \pi_2 L^2 - \ldots \equiv \phi(L)(1-L)^d$, and the π_j can be computed by using the binomial expansion of the fractional differencing operator $(1-L)^d$,

$$(1-L)^d = 1 - dL + \frac{d(d-1)L^2}{2!} - \frac{d(d-1)(d-2)L^3}{3!} + \cdots . \tag{13}$$

The AML estimator is asymptotically efficient if the errors ε_t are normally distributed. Under less restrictive regularity conditions, it is $\sqrt{n}$ consistent and asymptotically normal. To estimate the smooth transition ARFI model we modify Beran's estimator such that the π_j in (11) are now the autoregressive coefficients in the infinite order STAR representation of the model in (8) (see van Dijk, Franses and Paap, 2002, for further details).

We employ the Akaike Information Criterion (AIC) in the full ARFI-SCDXRL to select the appropriate autoregressive order p. The selected lag order $p = 2$ is subsequently imposed in the nested models, to facilitate comparison of the parameter estimates.

4 Estimation Results

All results discussed in this section are based on estimating models over the period from January 3, 1994 until December 31, 1997 (1011 observations). The remainder of the sample period will be used to evaluate the out-of-sample forecast performance of the various models. Detailed full-sample estimation results are available upon request. Table 3 contains estimation results for the different ARFI models in equations (4)-(7) which do not allow for structural change in μ_t ($\alpha_1 = \ldots = \alpha_5 = 0$). Table 4 shows results for the corresponding models which do allow for such level shifts in realized volatility. Several conclusions can be drawn from these tables. First, the order of integration d ranges between 0.3 and 0.5, which is in line with estimates reported in previous studies. For some models the point estimate of d is very close to 0.5, suggesting that log realized volatility may be non-stationary. Note however, that in all models, the autoregressive parameters ϕ_1 and ϕ_2 are negative, such that the model can be considered stationary for practical purposes such as forecasting at relatively short horizons. Comparing the estimates of d in Table 3 with those in Table 4 makes clear that allowing for structural change in μ_t lowers the order of integration, confirming that neglecting level shifts may spuriously suggest fractional integration, cf. Diebold and Inoue (2001). In the ARFI-DRL model, for example, d is estimated at 0.495, compared with 0.413 in the ARFI-SCDRL model. Note, however, that the point estimates of d are still significantly different from zero in the models with structural change. Hence, the level shift cannot fully account for the long memory feature in realized volatility. Also note that the order of integration is affected by the way the lagged returns are treated: if these are included as exogenous regressors, the estimate of d is substantially lower than if these are included in μ_t.

- insert Tables 3 and 4 about here -

Second, Figure 6 shows that the polynomial $P(t)$ captures the structural break at the end of 1996 quite accurately. The figure also shows that including a higher order polynomial is a more flexible approach compared to modeling the structural break using a logistic function.

- insert Figure 6 about here -

Third, the estimates of the seasonal dummy parameters $\delta_1, \ldots, \delta_5$ confirm the descriptive statistics in Table 2, in that on average realized volatility is significantly lower on Mondays and Tuesdays and higher on Fridays.

Fourth, the models that include lagged returns indicate a significant relationship between $\log(s_t^2)$ and r_{t-1}. We also find convincing evidence for the presence of a leverage effect. The point estimates of β_1 and β_3 in ARFI-(SC)D(X)RL models in fact suggest that only negative lagged returns affect current realized volatility, as $\hat{\beta}_1$ is not significantly different from zero.

Fifth, comparing the residual standard deviation, AIC and BIC across different columns shows that incorporating the different nonlinear features in the model enhances the in-sample fit. Allowing for a leverage effect appears to be most important in this respect, where the models which include the terms involving lagged returns as exogenous regressors (cf. Ebens, 1999) are preferred over models which include these terms in the conditional mean μ_t (cf. Oomen, 2002). Note that the AIC values for ARFI and ARFI-SC models do not differ substantially, suggesting that accounting for the level shift in realized volatility does not lead to much improvement of the model.

Finally, the results for the smooth transition model in Table 5 show that a large negative shock to volatility causes volatility to move to the lower regime ($G(s_t; \gamma, c) = 0$) where persistence is higher when compared to the upper regime ($G(s_t; \gamma, c) = 1$). In the infinite order STAR representation of the model $\widehat{\pi}_1$ equals $\widehat{d} + \widehat{\phi}_{2,1} = 0.117$ in the high regime and $\widehat{d} + \widehat{\phi}_{1,1} = 0.247$ in the low regime. The transition between the regimes occurs gradually, as shown by the point estimate of 2.845 for γ. We do not report standard errors for $\hat{\gamma}$ for reasons discussed extensively in Teräsvirta (1994, 1998) in the context of smooth transition (auto)regressions. Figure 7 shows the transition function against the transition variable and over time. It is evident that volatility is in the lower regime most of the time, but occasionally it moves towards the upper regime (For 5.9% of the observations the value of the transition function $G(s_t; \gamma, c)$ is larger than 0.5). Judging from the values for AIC and BIC, allowing for regime switching behaviour does not lead to an improvement in the in-sample fit.

- insert Table 5 and Figure 7 about here -

5 Forecasting Volatility

The period from January 2, 1998 through December 29, 2000 (756 observations) is used to judge the relevance of modeling the nonlinearities in S&P 500 realized volatility for out-of-sample forecasting purposes. All models are estimated recursively, using an expanding window of data. Volatility forecasts for 1- to 20-days ahead are constructed by means of Monte Carlo simulation, where we use the infinite order AR-representation of the ARFI models given in (12), truncated after 200 lags.[6] In addition to forecasts for logarithmic realized volatility, which is the

[6]The truncated infinite order AR representation in (12) can be rewritten as $y_t = \hat{\mu}_t + \sum_{j=1}^{p^*} \pi_j (y_{t-j} - \hat{\mu}_{t-j}) + e_t$, where $p^* = 200$. The 1-step ahead forecast $y_{t+1|t}$ is obtained by sampling B independent random draws $z_{t+1}^{(i)}$, $i = 1, \ldots, B$ from a standard normal distribution, which are multiplied by the residual standard deviation $\hat{\sigma}_e$. The resulting shocks $e_{t+1}^{(i)} = z_{t+1}^{(i)} \hat{\sigma}_e$ are fed into the model to obtain a realization $y_{t+1|t}^{(i)} = \hat{\mu}_{t+1} + \sum_{j=1}^{p^*} \pi_j (y_{t+1-j} - \hat{\mu}_{t+1-j}) + e_{t+1}^{(i)}$. Finally, the 1-step ahead forecast $y_{t+1|t}$ is the mean across all B realizations, $y_{t+1|t} = \frac{1}{B} \sum_{i=1}^{B} y_{t+1|t}^{(i)}$. For

dependent variable in the ARFI models, we also construct forecasts for the realized variance and realized standard deviation, where we ensure that these forecasts are unbiased.[7] In the forecast evaluation below, we concentrate on forecasts for the standard deviation. Results for forecasts of the (logarithmic) variance are qualitatively similar and are available upon request. Furthermore, h-days ahead forecasts refer to realized standard deviations over the next h days, i.e. $\widehat{s}_{t+h|t} = \sqrt{\sum_{j=1}^{h} \widehat{s}^2_{t+j|t}}$ (instead of daily realized standard deviation h-days ahead).

For comparison purposes, forecasts were also produced for a number of popular models that only use daily returns and treat volatility as a latent variable. First, Riskmetrics uses historical volatility with exponentially declining weights,

$$\sigma_t^2 = \lambda \sigma_{t-1}^2 + (1-\lambda) r_{t-1}^2, \tag{14}$$

with $\lambda = 0.94$.

Second, Glosten, Jagannathan and Runkle (GJR; 1993) incorporated leverage effects into the popular Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model. Also including day-of-week dummies, the GJR-GARCH(1,1)-D model is specified as

$$r_t = \mu + \varepsilon_t \tag{15}$$

$$\varepsilon_t | \Psi_{t-1} \sim N(0, \sigma_t^2) \tag{16}$$

$$\sigma_t^2 = \omega + \delta_1 D_{1,t}^* + \delta_2 D_{2,t}^* + \delta_4 D_{4,t}^* + \delta_5 D_{5,t}^* + \alpha \varepsilon_{t-1}^2 + \gamma \varepsilon_{t-1}^2 \mathrm{I}[\varepsilon_{t-1} < 0] + \beta \sigma_{t-1}^2, \tag{17}$$

where Ψ_{t-1} is the information set containing all daily information up to day t, and the error terms u_t are assumed to follow a conditional normal distribution with mean zero and variance σ_t^2. Restricting $\gamma = 0$ results in the GARCH(1,1)-D model, and further imposing $\delta_1 = \delta_2 = \delta_4 = \delta_5 = 0$ renders the standard GARCH(1,1) model. For the three GARCH models and Riskmetrics' approach forecasts are produced to compare with the ARFI models.

A number of out-of-sample forecast performance measures are used to evaluate and compare the various models. First, the quality of individual forecasts is assessed by regressing the observed h-day realized standard deviation on the corresponding

multiple step ahead forecasts from models which include lagged returns in μ_t, these returns are simulated as well, by multiplying the standard deviation in the i-th path by another draw v_t from a standard normal distribution, e.g. $r_{t+1}^{(i)} = \sqrt{\exp(y_{t+1|t}^{(i)})} v_t$.

[7] This is achieved by applying the appropriate transformation to all simulated paths of log realized volatility individually, and then averaging. For example, the 1-step ahead forecast of the realized standard deviation is computed as $s_{t+1|t} = \frac{1}{B} \sum_{i=1}^{B} \sqrt{\exp(y_{t+1|t}^{(i)})}$.

forecast,

$$s_{t+h|t+1} = \sqrt{\sum_{j=1}^{h} s_{t+j}^2} = b_0 + b_1 \widehat{s}_{t+h|t} + \nu_t, \tag{18}$$

where b_0 and b_1 should be equal to 0 and 1, respectively, for the forecast to be unbiased and efficient.

Forecasts from two different models are compared directly by means of the encompassing regression

$$s_{t+h|t+1} = \sqrt{\sum_{j=1}^{h} s_{t+j}^2} = b_0 + b_1 \widehat{s}_{t+h|t,(1)} + b_2 \widehat{s}_{t+h|t,(2)} + \nu_t, \tag{19}$$

where $\widehat{s}_{t+h|t,(i)}$ is the forecast of model $i = 1, 2$ for the volatility from $t + 1$ to $t + h$, made at the end of day t. Forecast encompassing is also tested by means of the statistics developed in Harvey, Leybourne and Newbold (1998). In addition a number of popular error metrics are computed, namely the Mean Squared Prediction Error (MSPE; $MSPE = \frac{1}{R}\sum_{i=1}^{R}(s_{t+h|t+1} - \widehat{s}_{t+h|t})^2$ where R denotes the number of forecasts), the Mean Absolute Error (MAE; $MAE = \frac{1}{R}\sum_{i=1}^{R}|s_{t+h|t+1} - \widehat{s}_{t+h|t}|$), and the Heteroskedasticity-adjusted MSPE (HMSPE; $HMSPE = \frac{1}{R}\sum_{i=1}^{R}(1 - \frac{\widehat{s}_{t+h|t}}{s_{t+h|t+1}})^2$). We employ Diebold-Mariano (1995) tests of equal forecast accuracy to assess whether differences in the error metrics of two competing models are significant. Finally, we also report the Mean Error (ME; $ME = \frac{1}{R}\sum_{i=1}^{R} s_{t+h|t+1} - \widehat{s}_{t+h|t}$).[8]

Results for the one-day ahead forecasts from the GARCH models and ARFI models which do not allow for structural change are presented in Table 6. We confirm the findings of earlier work, in that the ARFI models produce volatility forecasts that are superior to popular volatility models based on daily data. For example, the regression R^2 of the "basic" ARFI model is 42.1% compared to 31.0% for Riskmetrics' approach and 33.1% for the GARCH model. Including the leverage effect and day-of-the-week dummies, resulting in the GJR-GARCH(1,1)-D model, increases the regression R^2 to 39.8%, but this is still short of the R^2 of even the simplest ARFI model. The top panels in Figures 8 and 9 show the daily realized standard deviations for the out-of-sample period along with the one-day-ahead forecasts from the GARCH(1,1) model and the ARFI-DXRL model, respectively. It is seen that the GARCH(1,1) forecasts track the general pattern, or low-frequency movements in volatility quite well. The main advantage of the long memory model is that, in addition, it captures a much larger part of the high-frequency movements in volatility.

[8]See Andersen, Bollerslev and Meddahi (2003, 2004) for recent discussions on issues involved in evaluating realized volatility forecasts.

- insert Table 6 and Figures 8 and 9 about here -

The best forecast performance is attained by the ARFI-DXRL model, with a regression R^2 of 46.1%. The Diebold-Mariano statistics reported in the table compare the forecasts from this model with those of all other models (on a one-to-one basis). The statistics show that the MAE and HMSPE are significantly reduced from 0.264 to 0.248 and from 0.074 to 0.066, respectively, when comparing the nonlinear ARFI-DXRL model with the linear ARFI model. Including the leverage effect exogenously slightly improves over including the leverage effect in μ_t, the latter resulting in a regression R^2 of 45.2%. In general it is the leverage effect that contributes most of all nonlinearities to the improvement in forecast performance over the linear ARFI model. The encompassing regression results at the bottom of Table 6 show that the null that the forecasts from the ARFI-DXRL model encompass the forecasts from the Riskmetrics and GARCH models is rejected at conventional significance levels. Hence, the volatility forecasts from these models do have some incremental information not already contained in the nonlinear ARFI model. For example, the resulting regression R^2 of optimally (with hindsight) combining the GJR-GARCH model with the nonlinear ARFI model is 48.4%.

Table 7 summarizes the performance of the ARFI models which allow for structural change and level-dependent volatility persistence for one-day ahead forecasts. Comparing these results with the corresponding ones in Table 6, it is seen that modeling the apparent level shift in realized volatility does improve the out-of-sample forecast performance of the long memory models when judged by R^2. However, MSPE, MAE and HMSPE are somewhat larger, while only the ME is closer to 0 for the models with structural change. Allowing for different volatility regimes (STARFI-SCDXRL) does not lead to better out-of-sample forecasts.

- insert Table 7 about here -

The results for 5-day, 10-day and 20-day ahead forecasts of realized standard deviation are summarized in Table 8. As expected the benefits of modelling the nonlinearities explicitly become less for longer horizons. Comparing the ARFI and ARFI-DXRL models, the regression R^2 increases from 52.8% to 54.4% for the 5-day horizon, from 49.0% to 50.2% for the 10-day horizon, and even drops slightly from 37.8% to 37.7% for the 20-day horizon.

- insert Table 8 about here -

As an alternative method for evaluating the forecasts for the different models from a more economic point of view, we consider Value-at-Risk (VaR) estimates which are constructed using the volatility forecasts from the different models, see also Giot and Laurent (2004). Under the 1998 Market Risk Amendment (MRA) to the Basle Capital Accord, commercial banks are required to reserve capital to cover their market risk exposure. The required market risk capital for time $t+1$ (MRC_{t+1}) is determined by a bank's 99% VaR estimate calculated for a 10-day holding period (VaR_t^{10}) and is defined as the higher of the previous day's VaR estimate or the average of the estimates over the last sixty business days times a multiplication term[9]

$$MRC_{t+1} = \max\left(VaR_t^{10}, S_t \times \frac{1}{60} \sum_{i=0}^{59} VaR_{t-i}^{10} \right), \tag{20}$$

using the one-step ahead volatility forecasts from the different models.

Under the MRA, the accuracy of a bank's VaR estimates is judged by the accuracy of the 1-day ahead 99% VaR estimates[10]. We evaluate the accuracy of these estimates using the interval forecast evaluation techniques proposed by Christoffersen (1998) and the method set forth by Lopez (1999) which uses regulatory loss functions. The Christoffersen method is a test for conditional coverage, jointly testing the hypothesis that the percentage of times that the actual loss on a portfolio exceeds the VaR estimate (denoted by a VaR 'exception') equals one minus the significance level used in the VaR calculation (unconditional coverage) together with the hypothesis of serial independence for these VaR exceptions (independence).

Defining the indicator variable I_{t+1} as

$$I_{t+1} = \begin{cases} 1 & \text{if } r_{t+1} < VaR_t^1 \\ 0 & \text{if } r_{t+1} \geq VaR_t^1, \end{cases} \tag{21}$$

where r_{t+1} is the return over day $t+1$ and VaR_t^1 is the 100(1-α)% VaR estimate for day $t+1$ made on day t, a VaR exception corresponds with $I_{t+1} = 1$. The likelihood of observing x exceptions in a series of length R under the null hypothesis of accurate unconditional coverage is given by $L_0 = \alpha^x(1-\alpha)^{R-x}$. The corresponding likelihood under the alternative is $L_1 = \hat{\alpha}^x(1-\hat{\alpha})^{R-x}$, where $\hat{\alpha}$ is the "observed" probability of an exception, $\hat{\alpha} = x/R$. The null hypothesis of correct unconditional coverage can now be tested by means of the standard likelihood ratio statistic

$$LR_{uc} = 2[\log(L_1) - \log(L_0)], \tag{22}$$

[9] The value of the multiplication factor S_t is determined by the number of times that the actual losses on a portfolio exceed the 1-day 99% VaR estimates (so called VaR-exceptions). Three zones for an increasing number of exceptions are distinguished and the value of S_t increases from a minimum value of 3 to a possible maximum value of 4 across the different zones, see the Basle Committee on Banking Supervision (1996) for more details. In the evaluation of the VaR estimates we, however, fix S_t to the value 3. Note further that under the conditions of the MRA the VaR estimates are evaluated in dollar terms whereas we will consider the VaR in percentage terms.

[10] It is suggested by the Basle committee to use an evaluation period of at least 250 business days. Here we use the entire forecast evaluation sample.

which has an asymptotic $\chi^2(1)$ distribution.

The test statistic for the test of independence is also $\chi^2(1)$ distributed and is given by

$$LR_{ind} = 2[\log(L_2) - \log(L_1)], \tag{23}$$

where L_1 is the likelihood function under independence as given above, and L_2 is the likelihood function under the alternative of first-order Markov dependence. The latter is given by $L_2 = (1-\pi_{01})^{R_{00}}\pi_{01}^{R_{01}}(1-\pi_{11})^{R_{10}}\pi_{11}^{R_{11}}$ where R_{ij} is the number of observations with value i followed by value j.

Testing correct conditional coverage boils down to testing correct unconditional coverage and independence simultaneously. The likelihood functions under the null and alternative are given by L_0 and L_2, respectively and, hence, the likelihood ratio statistic for correct conditional coverage is simply the sum of the two previous statistics,

$$LR_{cc} = LR_{uc} + LR_{ind}, \tag{24}$$

which is asymptotically $\chi^2(2)$ distributed.

The alternative method proposed by Lopez (1999) for evaluating VaR forecasts is based on the use of loss functions that are more closely related to the regulatory VaR requirements. By choosing a specific loss function, one can assign a numerical score to each individual VaR estimate that reflects specific concerns that one may have. For example, it seems natural to not only consider the number of VaR exceptions but also the magnitude of exceptions since the latter can be quite substantial. Therefore, we consider the loss function[11]

$$C_{t+1} = \begin{cases} 1 + (r_{t+1} - VaR_t^1)^2 & \text{if } r_{t+1} < VaR_t^1 \\ 0 & \text{if } r_{t+1} \geq VaR_t^1, \end{cases} \tag{25}$$

which includes the squared magnitude of the VaR exception $(r_{t+1} - VaR_t^1)^2$. Given a sample of R VaR estimates the total loss for the sample is given by $C = \sum_{i=1}^{R} C_{t+1}$.

To assess the relative performance of each volatility forecasting model, we compute for each model the average and standard deviation of the capital requirement MRC_{t+1} over the forecast evaluation period. Furthermore, we determine the unconditional coverage $\hat{\alpha}$ together with the interval evaluation test statistics. Finally, we compute the total loss C for the sample of R one-day VaR estimates as well as the average score (defined as the total score divided by the number of exceptions) and the maximum daily score. The results are presented in Table 9.

[11]Lopez (1999) discusses several possible loss functions one of which is the binomial loss function given in (21). The loss function given in the main text is in line with the guidelines by the Basle Committee which states that both the number as well the magnitude of exceptions are a matter of concern to regulators.

- insert Table 9 about here -

We first of all observe that the average capital requirement is comparable across the different models. However, the long memory models typically have considerable less fluctuation in the required level of capital. This is confirmed graphically by Figure 10, showing how the capital requirement evolves over time for the Riskmetrics, GJR-GARCH(1,1)-D and ARFI-DXRL models.

- insert Figure 10 about here -

All models have a higher unconditional coverage than expected, as judged from the estimated $\hat{\alpha}$, leading to strong rejections of the null of correct unconditional coverage in all cases. This might be due to our use of the normal distribution for the standardized shocks ε/σ_t. Giot and Laurent (2004) obtain more accurate unconditional coverage using a (skewed) Student t distribution, both for GARCH and for realized volatility models. By contrast, the null of independence is not rejected for any models under consideration. Based on the quadratic magnitude loss function, the ARFI type models again perform well when compared to the GARCH type models. Among the latter class of models the GJR-GARCH(1,1)-D performs best, although still slightly worse than the ARFI-DXRL model. The 1-day VaR estimates from these two models actually are quite different, as shown in Figure 11.

- insert Figure 11 about here -

6 Foreign exchange rates

To check the robustness of the conclusions for the S&P 500, and to test the adequacy of the proposed nonlinear model for other realized volatility series, we also consider the DM/\$, ¥/\$ and ¥/DM exchange rates. Realized volatilities are constructed by summing 288 5-minute squared returns, with the sample period running from December 1, 1986 until June 30, 1999. The characteristics of these data are described in detail in Andersen, Bollerslev, Diebold and Labys (2001). All models are estimated for an in-sample period from December 1, 1986 until November 30, 1996 (2449 observations), and out-of-sample forecasts are produced for the remaining 596 observations. No significant level shifts were detected in the exchange rate volatilities. To save space, we only summarize the in-sample fit and the out-of-sample forecast performance in Table 10.[12]

- insert Table 10 about here -

[12]Model estimates and graphs in the spirit of the S&P 500 results are available upon request.

The in-sample results for the ARFI models for the three exchange rates show a pattern similar to the S&P 500. In all cases the in-sample fit is in general improved when accounting for the various nonlinearities present in the data, despite the penalties for adding additional parameters.

Even more so than for the S&P 500, the out-of-sample forecast performance of the ARFI models is found to be superior to that of standard models based on daily returns only. For the ¥/$, for example, the R^2 for the regression of the daily realized volatility on its one-day ahead forecast increases from 25.6% for the GARCH(1,1) model to 36.6% for the ARFI model.[13] Adding daily dummies (ARFI-D) increases the R^2 to 40.3%, while the best way to incorporate the lagged return effects is by including them as exogeneous variables. Adding the lagged absolute return (ARFI-DXR) increases the R^2 to 49.0%, and even to 55.5% when also including the leverage effect (ARFI-DXRL). For all exchange rates the introduction of different AR parameters as a function of the level of the volatility does not improve the out-of-sample forecast performance, but it also does not worsen it.

7 Concluding Remarks

In this paper we propose a nonlinear long-memory time series model for realized volatility that incorporates all well-known stylized facts from the (GARCH) volatility literature, in particular level shifts, day-of-the-week effects, leverage effects and volatility level effects. The model, as well as several restricted versions, are estimated for the S&P 500 index and three exchange rates.

The in-sample results show that all nonlinearities are highly significant and improve the description of the data. The out-of-sample results show that for shorter horizons, up to 10 days, accounting for these nonlinearities significantly improves the forecast performance compared to a linear ARFI model. Such short-term volatility forecasts are especially useful for short-term risk management, including Value-at-Risk. For longer horizons no benefit is obtained from incorporating nonlinearities.

The most important nonlinearities are the leverage effect for the S&P 500 index, and the leverage effect as well as the day-of-the-week effects for the exchange rates. The best way to incorporate the effects of lagged daily returns is to include them as exogenous regressors, i.e. outside the long memory filter. Not important for the forecast performance is allowing the persistence of shocks to depend on the level of volatility, and modeling the level shifts for the S&P 500 index.

[13] For the same data and out-of-sample period but using realized volatilities based on 30-minute returns (instead of 5-minute returns), Andersen, Bollerslev, Diebold and Labys (2003) find an R^2 of 32.9%.

References

Aït-Sahalia, Y., P. Mykland and L. Zhang (2004), How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise, *Review of Financial Studies*, in press.

Andersen, T.G. and T. Bollerslev (1998), Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts, *International Economic Review* **39**, 885–905.

Andersen, T.G., T. Bollerslev and F.X. Diebold (2002), Parametric and Nonparametric volatility measurement, forthcoming in Aït-Sahalia and L.P. Hansen (eds.), *Handbook of Financial Econometrics*, Amsterdam: North Holland.

Andersen, T.G., T. Bollerslev, F.X. Diebold and H. Ebens (2001), The Distribution of Realized Stock Return Volatility, *Journal of Financial Economics* **61**, 43–76.

Andersen T.G., T. Bollerslev, F.X. Diebold and P. Labys (2001), The Distribution of Realized Exchange Rate Volatility, *Journal of the American Statistical Association* **96**, 42–55.

Andersen, T.G., T. Bollerslev and N. Meddahi (2003), Correcting the Errors: A Note on Volatility Forecast Evaluation based on High-Frequency Data and Realized Volatilities, Working Paper No. 322, Kellog School of Management, Northwestern University.

Andersen, T.G., T. Bollerslev and N. Meddahi (2004), Analytic Evaluation of Volatility Forecasts, *International Economic Review*, in press.

Andersen, T.G., T. Bollerslev, F.X. Diebold and P. Labys (2003), Modeling and Forecasting Realized Volatility, *Econometrica* **71**, 579–625.

Andreou E. and E. Ghysels (2002), Detecting Multiple Breaks in Financial Market Volatility Dynamics, *Journal of Applied Econometrics* **17**, 579–600.

Areal, N.M.P.C. and S.J. Taylor (2002), The Realized Volatility of FTSE-100 Futures Prices, *Journal of Futures Markets* **22**, 627–48.

Baillie, R. and T. Bollerslev (1989), The Message in Daily Exchange Rates: A Conditional Variance Tale, *Journal of Business & Economic Statistics* **7**, 279–305.

Bandi, F.M. and J.R. Russell (2003), Microstructure Noise, Realized Volatility, and Optimal Sampling, working paper, University of Chicago.

Bandi, F.M. and J.R. Russell (2004), Separating Microstructure Noise from Volatility, working paper, University of Chicago.

Barndorff-Nielsen, O.E. and N. Shephard (2002a), Econometric Analysis of Realised Volatility and its use in Estimating Stochastic Volatility Models, *Journal of the Royal Statistical Society* **B 64**, 253–280.

Barndorff-Nielsen, O.E. and N. Shephard (2002b), Estimating Quadratic Variation Using Realized Variance, *Journal of Applied Econometrics* **17**, 457–478.

Barndorff-Nielsen, O.E. and N. Shephard (2003), Realised Power Variation and Stochastic Volatility, *Bernoulli* **9**, 243–265 and 1109–1111.

Barndorff-Nielsen, O.E. and N. Shephard (2004a), Power and Bipower Variation with

Stochastic Volatility and Jumps, *Journal of Financial Econometrics* **2**, 1–48.

Barndorff-Nielsen, O.E. and N. Shephard (2004b), How Accurate is the Asymptotic Approximation to the Distribution of Realised Volatility, in D.A. Andrews, J. Powell, P.A. Ruud, and J.H. Stock (eds.), *Identification and Inference for Econometric Models. A Festschrift in Honour of T.J. Rothenberg*, Cambridge: Cambridge University Press, in press.

Basle Committee on Banking Supervision (1996), *Supervisory Framework for the Use of Backtesting in Conjunction with the Internal Models Approach to Market Risk Capital Requirements*, Bank for International Settlements Publication No. 22.

Beran, J. (1995), Maximum Likelihood Estimation of the Differencing Parameter for Invertible Short and Long Memory Autoregressive Integrated Moving Average Models, *Journal of the Royal Statistical Society* **B 57**, 659–672.

Black, F. (1976), Studies of Stock Market Volatility Changes, *Proceedings of the American Statistical Association, Buinsess and Economic Statistics Section*, 177–181.

Bollerslev T. and J.H. Wright (2000), Semiparametric estimation of long-memory volatility dependencies: The role of high-frequency data, *Journal of Econometrics* **98**, 81–106.

Christensen, B.J. and N.R. Prabhala (1998), The Relation Between Implied and Realized Volatility, *Journal of Financial Economics* **50**, 125–150.

Christoffersen, P.F. (1998), Evaluating Interval Forecasts, *International Economic Review* **39**, 841–862.

Diebold, F.X., and A. Inoue (2001), Long Memory and Regime Switching, *Journal of Econometrics* **105**, 131–159.

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

Ebens, H. (1999), Realized Stock Volatility, unpublished manuscript, Johns Hopkins University.

Engle, R.F. and V.K. Ng (1993), Measuring and Testing the Impact of News on Volatility, *Journal of Finance* **48**, 1749–1778.

Giot, P. and S. Laurent (2004), Modelling Daily Value-at-Risk Using Realized Volatility and ARCH Type Models, *Journal of Empirical Finance*, in press.

Glosten, L.R., R. Jagannathan and D.E. Runkle (1993), On the Relation Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks, *Journal of Finance* **48**, 1779–1801.

Harvey, C.R. and R.D. Huang (1991), Volatility in the Foreign Currency Futures Market, *Review of Financial Studies* **4**, 543–569.

Harvey, D.I., S.J. Leybourne and P. Newbold (1998), Tests for Forecast Encompassing, *Journal of Business & Economic Statistics* **16**, 254–259.

Hol, E., B. Jungbacker and S.J. Koopman (2004), Forecasting Daily Variability of the S&P 100 Stock Index using Historical, Realised and Implied Volatility Measurements, Discussion Paper No. 2004-016/4, Tinbergen Institute.

Jorion, P. (1995), Predicting Volatility in the Foreign Exchange Market, *Journal of Finance* **50**, 507–528.

Lamoureux, C.G. and W.D. Lastrapes (1990), Persistence in Variance, Structural Change, and the GARCH Model, *Journal of Business & Economic Statistics* **8**, 225–234.

Li, K. (2002), Long Memory Versus Option-Implied Volatility Predictions, *Journal of Derivatives* **9**,9–25.

Longin, F.M. (1997), The Threshold Effect in Expected Volatility: A Model Based on Asymmetric Information, *Review of Financial Studies* **10**, 837–869.

Lopez, J.A. (1999), Methods for Evaluating Value-at-Risk Estimates, *Federal Reserve Bank of San Francisco* **2**, 3–17.

Maheu, J.M. and T.H. McCurdy (2002), Nonlinear Features of Realized FX Volatility, *Review of Economics and Statistics* **84** 668–681.

Martens, M. (2002), Measuring and Forecasting S&P 500 Index-Futures Volatility Using High-Frequency Data, *Journal of Futures Markets* **22**, 497–518.

Martens, M. and J. Zein (2004), Forecasting Financial Volatility: High-Frequency Time-Series Forecasts Vis--Vis Implied Volatility, *Journal of Futures Markets*, in press.

Merton, R.C. (1980), On Estimating the Expected Return on the Market; An Exploratory Investigation, *Journal of Financial Economics* **8**, 323–361.

Oomen, R.C.A. (2002), Modelling Realized Variance when Returns are Serially Correlated, working paper, University of Warwick.

Pagan, A.R. and G.W. Schwert (1990), Alternative Models for Conditional Stock Volatility, *Journal of Econometrics* **45**, 267–290.

Pong, S., M.B. Shackleton, S.J. Taylor and X. Xu (2004), Forecasting Currency Volatility: A Comparison of Implied Volatilities and AR(FI)MA Models, *Journal of Banking and Finance*, in press.

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

Teräsvirta, T. (1998), Modelling economic relationships with smooth transition regressions, in A. Ullah and D.E.A. Giles (eds.), *Handbook of Applied Economic Statistics*, New York: Marcel Dekker, pp. 507–552.

Thomakos, D.D. and T. Wang (2003), Realized Volatility in the Futures Markets, *Journal of Empirical Finance* **10**, 321–353.

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

Zhang, L., P. Mykland and Y. Aït-Sahalia (2003), A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data, working paper, Princeton University.

Table 1: Descriptive Statistics for Daily S&P 500 Return and Realized Volatility

	Mean	Med	Min	Max	Std.dev.	Skew	Kurt
Returns	0.043	0.066	−7.811	5.671	1.102	−0.365	8.175
Squared returns	1.217	0.320	0	61.005	3.242	9.835	145.260
Absolute returns	0.978	0.709	0	9.789	0.977	2.583	15.142
Standardized returns	0.124	0.083	−2.623	3.025	1.000	0.104	2.762
Realized variance	1.194	0.729	0.065	32.996	1.770	7.250	89.405
Realized std. deviation	0.962	0.853	0.256	5.744	0.516	2.356	13.642
Log realized variance	−0.308	−0.316	−2.727	3.496	0.937	0.295	3.025

Notes: The table contains summary statistics for daily S&P 500 return and realized volatility measures. The sample period covers January 3, 1994 until December 29, 2000 (1767 observations). Absolute returns are multiplied with $\sqrt{\pi/2}$. Standardized returns are obtained by dividing the raw returns by the realized standard deviation.

Table 2: Day-of-the-Week Effects in S&P 500 Return and Realized Volatility

	Overall	MON	TUE	WED	THU	FRI
Returns	0.043	0.081	0.085	−0.018	−0.021	0.093
Squared returns	1.217	1.298	1.362	0.865	1.219	1.348
Absolute returns	0.978	0.954	1.025	0.869	1.004	1.040
Standardized returns	0.124	0.181	0.129	0.064	0.079	0.173
Realized variance	1.191	1.073	1.158	1.093	1.306	1.324
Realized std. deviation	0.962	0.907	0.942	0.951	0.986	1.020
Log realized variance	−0.308	−0.430	−0.352	−0.295	−0.278	−0.191

Notes: The table contains daily means for S&P 500 returns and realized volatility measures. The sample period covers January 3, 1994 until December 29, 2000 (1767 observations). Absolute returns are multiplied with $\sqrt{\pi/2}$. Standardized returns are obtained by dividing the raw returns by the realized standard deviation.

Table 3: Estimated ARFI models for daily S&P 500 realized volatility, January 1994-December 1997

	ARFI	ARFI-D	ARFI-DR	ARFI-DRL	ARFI-DXR	ARFI-DXRL
$\hat{d}$	0.471	0.477	0.471	0.495	0.363	0.386
	(0.046)	(0.047)	(0.046)	(0.045)	(0.036)	(0.037)
$\widehat{\alpha_0}$	−0.785	−0.791	−0.876	−0.877	−1.652	−1.539
	(0.311)	(0.320)	(0.299)	(0.316)	(0.214)	(0.264)
$\widehat{\beta_1}$	–	–	0.149	0.006	0.185	0.016
			(0.038)	(0.044)	(0.037)	(0.039)
$\widehat{\beta_2}$	–	–	–	−0.053	–	−0.033
				(0.046)		(0.045)
$\widehat{\beta_3}$	–	–	–	0.319	–	0.357
				(0.053)		(0.050)
$\widehat{\delta_1}$	–	−0.136	−0.142	−0.141	−0.141	−0.140
		(0.033)	(0.033)	(0.032)	(0.034)	(0.032)
$\widehat{\delta_2}$	–	−0.065	−0.070	−0.072	−0.071	−0.074
		(0.031)	(0.030)	(0.030)	(0.030)	(0.030)
$\widehat{\delta_4}$	–	−0.001	0.013	0.014	0.013	0.016
		(0.031)	(0.031)	(0.030)	(0.031)	(0.030)
$\widehat{\delta_5}$	–	0.171	0.173	0.158	0.176	0.158
		(0.033)	(0.033)	(0.033)	(0.033)	(0.033)
$\widehat{\phi_1}$	−0.104	−0.089	−0.149	−0.249	−0.058	−0.157
	(0.053)	(0.054)	(0.056)	(0.057)	(0.049)	(0.048)
$\widehat{\phi_2}$	−0.088	−0.089	−0.076	−0.086	−0.044	−0.046
	(0.041)	(0.041)	(0.042)	(0.044)	(0.039)	(0.039)
$\hat{\sigma}_\varepsilon$	0.581	0.568	0.562	0.548	0.558	0.540
AIC	−1.078	−1.114	−1.133	−1.179	−1.149	−1.209
BIC	−1.058	−1.075	−1.089	−1.126	−1.105	−1.155
$\mathrm{LM_{SC}}(1)$	0.121	0.007	0.147	0.350	2.40	11.53
	(0.73)	(0.93)	(0.71)	(0.55)	(0.12)	(7.1E−4)
$\mathrm{LM_{SC}}(5)$	0.768	0.821	0.630	0.914	0.716	3.27
	(0.57)	(0.53)	(0.68)	(0.47)	(0.61)	(6.2E−3)
$\mathrm{LM_{SC}}(10)$	0.883	0.494	0.414	0.524	0.673	2.43
	(0.55)	(0.89)	(0.94)	(0.87)	(0.75)	(7.3E−3)

Notes: The table presents parameter estimates, diagnostic measures and misspecification tests for estimated ARFI models for daily S&P 500 realized volatility, over the period January 3, 1994-December 31, 1997 (1011 observations). The column ARFI-D(X)RL contains estimates of the model including daily dummies and leverage effects as part of μ_t (as exogenous variables), the column ARFI-D(X)R contains estimates of the model including daily dummies and symmetric effects of the lagged absolute return as part of μ_t (as exogenous variable) ($\beta_2 = \beta_3 = 0$), the column ARFI-D estimates for a model with daily dummies but without the lagged absolute return ($\beta_1 = \beta_2 = \beta_3 = 0$), and the column ARFI shows estimates for a model without daily dummies and without lagged absolute returns ($\beta_1 = \beta_2 = \beta_3 = 0$ and $\delta_1 = \ldots = \delta_5 = 0$). $\hat{\sigma}_\varepsilon$ is the residual standard deviation. $\mathrm{LM_{SC}}(q)$ denotes the (F variant of the) LM test of no serial correlation in the residuals up to and including order q. The numbers in parentheses below parameter estimates and test statistics are heteroskedasticity-consistent standard errors and p-values, respectively.

Table 4: Estimated ARFI models for daily S&P 500 realized volatility, January 1994-December 1997

	ARFI-SC	ARFI-SCD	ARFI-SCDR	ARFI-SCDRL	ARFI-SCDXR	ARFI-SCDXRL
$\hat{d}$	0.377	0.384	0.378	0.413	0.323	0.344
	(0.052)	(0.053)	(0.053)	(0.050)	(0.048)	(0.044)
$\widehat{\alpha_0}$	−1.318	−1.328	−1.386	−1.426	−1.443	−1.391
	(0.473)	(0.482)	(0.448)	(0.454)	(0.381)	(0.375)
$\widehat{\beta_1}$	–	–	0.151	0.007	0.183	0.003
			(0.039)	(0.044)	(0.039)	(0.042)
$\widehat{\beta_2}$	–	–	–	−0.053	–	−0.044
				(0.046)		(0.049)
$\widehat{\beta_3}$	–	–	–	0.318	–	0.367
				(0.053)		(0.053)
$\widehat{\delta_1}$	–	−0.136	−0.142	−0.141	−0.141	−0.141
		(0.033)	(0.034)	(0.033)	(0.034)	(0.032)
$\widehat{\delta_2}$	–	−0.064	−0.070	−0.072	−0.071	−0.073
		(0.031)	(0.030)	(0.030)	(0.030)	(0.030)
$\widehat{\delta_4}$	–	−0.001	0.013	0.014	0.013	0.016
		(0.031)	(0.031)	(0.030)	(0.031)	(0.030)
$\widehat{\delta_5}$	–	0.171	0.173	0.158	0.176	0.158
		(0.033)	(0.033)	(0.033)	(0.033)	(0.033)
$\widehat{\phi_1}$	−0.023	−0.007	−0.068	−0.178	−0.022	−0.118
	(0.059)	(0.059)	(0.060)	(0.060)	(0.056)	(0.053)
$\widehat{\phi_2}$	−0.049	−0.051	−0.035	−0.046	−0.026	−0.023
	(0.042)	(0.042)	(0.043)	(0.045)	(0.041)	(0.041)
$\hat{\sigma}_\varepsilon$	0.578	0.566	0.559	0.546	0.557	0.539
AIC	−1.078	−1.114	−1.134	−1.179	−1.143	−1.203
BIC	−1.034	−1.051	−1.065	−1.101	−1.074	−1.125
$\text{LM}_{\text{SC}}(1)$	0.355	0.708	0.339	0.141	2.847	11.620
	(0.552)	(0.400)	(0.560)	(0.707)	(0.092)	(6.7E−4)
$\text{LM}_{\text{SC}}(5)$	1.155	1.170	0.842	1.215	0.838	3.497
	(0.323)	(0.322)	(0.520)	(0.300)	(0.523)	(3.9E−3)
$\text{LM}_{\text{SC}}(10)$	1.190	0.882	0.691	0.766	0.857	2.658
	(0.294)	(0.550)	(0.733)	(0.662)	(0.573)	(3.3E−3)

Notes: The table presents parameter estimates, diagnostic measures and misspecification tests for estimated ARFI models for daily S&P 500 realized volatility, over the period January 3, 1994-December 31, 1997 (1011 observations). The column ARFI-SCD(X)RL contains estimates of the full model including structural change, daily dummies and leverage effects as part of μ_t (as exogenous variables), the column ARFI-SCD(X)R contains estimates of the model including structural change, daily dummies and symmetric effects of the lagged absolute return as part of μ_t (as exogenous variable) only ($\beta_2 = \beta_3 = 0$), the column ARFI-SCD estimates for a model with structural change and daily dummies ($\beta_1 = \beta_2 = \beta_3 = 0$), and the column ARFI-SC shows estimates for a model with structural change but without daily dummies and without lagged absolute returns ($\beta_1 = \beta_2 = \beta_3 = 0$ and $\delta_1 = \ldots = \delta_5 = 0$). $\hat{\sigma}_\varepsilon$ is the residual standard deviation. $\text{LM}_{\text{SC}}(q)$ denotes the (F variant of the) LM test of no serial correlation in the residuals up to and including order q. The numbers in parentheses below parameter estimates and test statistics are heteroskedasticity-consistent standard errors and p-values, respectively.

Table 5: Estimated Smooth Transition ARFI models for daily S&P 500 realized volatility, January 1994-December 1997

	STARFI-SCDXRL
$\hat{d}$	0.324
	(0.043)
$\widehat{\gamma}$	2.845
	-
$\widehat{c}$	0.016
	(0.533)
$\widehat{\alpha_0}$	-1.517
	(0.336)
$\widehat{\beta_1}$	-0.028
	(0.043)
$\widehat{\beta_2}$	-0.078
	(0.049)
$\widehat{\beta_3}$	0.415
	(0.055)
$\widehat{\delta}_1$	-0.137
	(0.032)
$\widehat{\delta}_2$	-0.078
	(0.030)
$\widehat{\delta}_4$	0.021
	(0.029)
$\widehat{\delta}_5$	0.156
	(0.033)
$\widehat{\phi}_{1,1}$	-0.077
	(0.059)
$\widehat{\phi}_{1,2}$	-0.057
	(0.054)
$\widehat{\phi}_{2,1}$	-0.207
	(0.108)
$\widehat{\phi}_{2,2}$	0.272
	(0.200)
$\hat{\sigma}_\varepsilon$	0.538
AIC	-1.202
BIC	-1.104
$\mathrm{LM}_{\mathrm{SC}}(1)$	5.515
	(0.019)
$\mathrm{LM}_{\mathrm{SC}}(5)$	1.729
	(0.125)
$\mathrm{LM}_{\mathrm{SC}}(10)$	1.825
	(0.052)

Notes: The table presents parameter estimates, diagnostic measures and misspecification tests for a Smooth Transition ARFI model for daily S&P 500 realized volatility, over the period January 3, 1994-December 31, 1997 (1011 observations). The column STARFI-SCDXRL contains estimates of the full model including regime-dependent volatility persistence, structural change and daily dummies as part of μ_t and leverage effects as exogenous variables. See Table 4 for further details.

Table 6: Out-of-Sample Forecast Evaluation, January 1998-December 2000 - standard deviation, one-day ahead

	b_0	b_1	b_2	R^2	ME	MSPE	MAE	HMSPE
Riskmetrics	0.267	0.750		0.310	−0.051	0.216	0.328	0.131
	(0.069)	(0.060)			(0.017)	[−4.78]	[−9.26]	[−8.44]
GARCH(1,1)	0.323	0.711		0.331	−0.044	0.215	0.312	0.121
	(0.060)	(0.053)			(0.017)	[−4.02]	[−6.83]	[−6.79]
GARCH-D(1,1)	0.339	0.705		0.328	−0.032	0.215	0.309	0.121
	(0.059)	(0.052)			(0.017)	[−3.98]	[−6.50]	[−6.41]
GJR-G-D(1,1)	0.353	0.721		0.398	0.017	0.195	0.284	0.091
	(0.049)	(0.045)			(0.016)	[−2.41]	[−3.91]	[−4.80]
ARFI	−0.037	1.059		0.421	0.033	0.172	0.264	0.074
	(0.082)	(0.076)			(0.015)	[−1.21]	[−2.88]	[−2.47]
ARFI-D	0.006	1.021		0.417	0.032	0.173	0.264	0.073
	(0.084)	(0.078)			(0.015)	[−1.37]	[−3.08]	[−2.50]
ARFI-DR	0.066	0.970		0.411	0.030	0.174	0.264	0.077
	(0.090)	(0.082)			(0.015)	[−1.15]	[−2.82]	[−2.66]
ARFI-DRL	0.133	0.911		0.452	0.026	0.163	0.251	0.069
	(0.072)	(0.065)			(0.015)	[−0.01]	[−1.30]	[−2.33]
ARFI-DXR	0.094	0.942		0.405	0.024	0.176	0.265	0.079
	(0.090)	(0.081)			(0.015)	[−1.25]	[−2.94]	[−2.90]
ARFI-DXRL	0.196	0.848		0.461	0.012	0.163	0.248	0.066
	(0.078)	(0.069)			(0.015)			
ARFI-DXRL +								
Riskmetrics	0.088	0.707	0.219	0.475				
	(0.068)	(0.076)	(0.072)	[2.30]				
GARCH(1,1)	0.120	0.696	0.206	0.474				
	(0.064)	(0.084)	(0.065)	[2.48]				
GARCH-D(1,1)	0.126	0.700	0.198	0.473				
	(0.064)	(0.085)	(0.064)	[2.46]				
GJR-G-D(1,1)	0.152	0.603	0.283	0.484				
	(0.063)	(0.091)	(0.066)	[3.33]				

Notes: The table presents estimates of regressions of realized standard deviation for the S&P 500 on a constant and one-day ahead forecasts from different models. The regression is $s_{t+1} = b_0 + b_1\widehat{s}_{t+1|t,(1)}\ [+b_2\widehat{s}_{t+1|t,(2)}\,] + u_t$, where $\widehat{s}_{t+1|t,(i)}$ is the one-day ahead forecast of the realized standard deviation from model i. The forecast evaluation period covers January 2, 1998-December 29, 2000 ($R = 756$). ARFI-D(X)RL refers to the full model including daily dummies and leverage effects as part of μ_t (as exogenous variables), ARFI-DR refers to the model including daily dummies and symmetric effects of the lagged absolute return as part of μ_t (as exogenous variables) ($\beta_2 = \beta_3 = 0$), ARFI-D to the model with daily dummies but without the lagged absolute return ($\beta_1 = \beta_2 = \beta_3 = 0$), and ARFI to the model without daily dummies and without lagged absolute returns ($\beta_1 = \beta_2 = \beta_3 = 0$ and $\delta_1 = \ldots = \delta_5 = 0$). Figures in brackets below b_j, $j = 0, 1, 2$ and ME are heteroskedasticity and autocorrelation-consistent standard errors. Figures in straight brackets below MSPE, MAE and HMSPE are Diebold-Mariano statistics of equal forecast accuracy, comparing the relevant model with the ARFI-DXRL model, where negative values indicate that the ARFI-DXRL model is more accurate. Figures in straight brackets below the R^2 in the encompassing regressions are Diebold-Mariano type forecasting encompassing tests, testing the null that the forecasts from the ARFI-DXRL model encompass the forecasts from the alternative model that is included in the regression.

Table 7: Out-of-Sample Forecast Evaluation, January 1998-December 2000 - standard deviation, one-day ahead

	b_0	b_1	b_2	R^2	ME	MSPE	MAE	HMSPE
ARFI-SC	0.117	0.896		0.416	−0.012	0.174	0.276	0.087
	(0.077)	(0.069)			(0.015)	[−1.14]	[−4.35]	[−4.97]
ARFI-SCD	0.147	0.871		0.412	−0.013	0.176	0.276	0.086
	(0.079)	(0.071)			(0.015)	[−1.32]	[−4.69]	[−4.68]
ARFI-SCDR	0.192	0.833		0.407	−0.015	0.179	0.275	0.090
	(0.088)	(0.077)			(0.015)	[−1.21]	[−4.18]	[−3.89]
ARFI-SCDRL	0.227	0.803		0.454	−0.018	0.169	0.263	0.080
	(0.062)	(0.055)			(0.015)	[−1.13]	[−5.26]	[−6.49]
ARFI-SCDXR	0.226	0.827		0.407	0.017	0.180	0.270	0.083
	(0.082)	(0.074)			(0.015)	[−1.32]	[−3.28]	[−2.90]
ARFI-SCDXRL	0.272	0.782		0.466	0.006	0.168	0.253	0.069
	(0.066)	(0.059)			(0.015)	[−1.46]	[−2.87]	[−2.49]
STARFI-SCDXRL	0.311	0.740		0.455	−0.010	0.177	0.259	0.075
	(0.084)	(0.073)			(0.015)	[−1.33]	[−3.03]	[−3.69]

Notes: The table presents estimates of regressions of realized standard deviation for the S&P 500 on a constant and one-day ahead forecasts from different models. The regression is $s_{t+1} = b_0 + b_1 \widehat{s}_{t+1|t,(1)} + u_t$, where $\widehat{s}_{t+1|t,(1)}$ is the one-day ahead forecast of the realized standard deviation from model i. The forecast evaluation period covers January 2, 1998-December 29, 2000 ($R = 756$). [ST]ARFI-SCD(X)RL refers to the full model including [regime-dependent volatility persistence], structural change, daily dummies and leverage effects as part of μ_t (as exogenous variables), ARFI-SCDR refers to the model including structural change, daily dummies and symmetric effects of the lagged absolute return as part of μ_t (as exogenous variables) ($\beta_2 = \beta_3 = 0$), ARFI-SCD to the model with structural change, daily dummies but without the lagged absolute return ($\beta_1 = \beta_2 = \beta_3 = 0$), and ARFI-SC to the model with structural change but without daily dummies and without lagged absolute returns ($\beta_1 = \beta_2 = \beta_3 = 0$ and $\delta_1 = \ldots = \delta_5 = 0$). Figures in brackets below b_j, $j = 0, 1$ and ME are heteroskedasticity and autocorrelation-consistent standard errors. Figures in straight brackets below MSPE, MAE and HMSPE are Diebold-Mariano statistics of equal forecast accuracy, comparing the relevant model with the ARFI-DXRL model, where negative values indicate that the ARFI-DXRL model is more accurate.

Table 8: Out-of-Sample Forecast Evaluation, January 1998-December 2000 - standard deviation, five, ten and twenty-days ahead

	5-days ahead					10-days ahead					20-days ahead				
	b_0	b_1	b_2	R^2	MSPE	b_0	b_1	b_2	R^2	MSPE	b_0	b_1	b_2	R^2	MSPE
Riskmetrics	0.327	0.733		0.408	0.137	0.390	0.693		0.410	0.126	0.490	0.623		0.381	0.124
	(0.089)	(0.075)			[−4.13]	(0.105)	(0.089)			[−2.70]	(0.125)	(0.105)			[−1.30]
GARCH(1,1)	0.387	0.691		0.416	0.141	0.428	0.668		0.420	0.128	0.471	0.646		0.413	0.117
	(0.070)	(0.060)			[−3.82]	(0.090)	(0.077)			[−2.63]	(0.104)	(0.083)			[−0.94]
GARCH-D(1,1)	0.401	0.687		0.408	0.144	0.439	0.669		0.410	0.130	0.472	0.656		0.403	0.118
	(0.068)	(0.059)			[−3.71]	(0.089)	(0.077)			[−2.71]	(0.107)	(0.086)			[−1.02]
GJR-G-D(1,1)	0.453	0.675		0.434	0.146	0.502	0.651		0.401	0.142	0.521	0.658		0.376	0.135
	(0.063)	(0.056)			[−3.23]	(0.089)	(0.079)			[−2.38]	(0.117)	(0.104)			[−1.46]
ARFI	−0.107	1.133		0.528	0.105	−0.094	1.133		0.490	0.102	0.015	1.068		0.378	0.111
	(0.101)	(0.091)			[−1.81]	(0.140)	(0.125)			[−1.16]	(0.218)	(0.197)			[−0.61]
ARFI-D	−0.099	1.125		0.531	0.104	−0.089	1.129		0.494	0.100	0.018	1.065		0.382	0.110
	(0.099)	(0.088)			[−1.68]	(0.137)	(0.123)			[−1.05]	(0.215)	(0.195)			[−0.55]
ARFI-DR	−0.084	1.113		0.527	0.104	−0.080	1.123		0.489	0.101	0.030	1.056		0.373	0.112
	(0.101)	(0.090)			[−2.07]	(0.142)	(0.128)			[−1.18]	(0.222)	(0.202)			[−0.67]
ARFI-DRL	−0.037	1.064		0.533	0.101	−0.030	1.072		0.491	0.099	0.084	1.000		0.374	0.109
	(0.091)	(0.081)			[−1.27]	(0.139)	(0.124)			[−1.12]	(0.221)	(0.199)			[−0.63]
ARFI-DXR	−0.039	1.065		0.515	0.105	−0.015	1.055		0.475	0.101	0.139	0.946		0.352	0.112
	(0.099)	(0.083)			[−2.86]	(0.139)	(0.122)			[−1.77]	(0.214)	(0.189)			[−1.03]
ARFI-DXRL	0.049	0.973		0.544	0.097	0.057	0.975		0.502	0.094	0.196	0.873		0.377	0.105
	(0.077)	(0.065)				(0.108)	(0.092)				(0.176)	(0.152)			
ARFI-DXRL+															
Riskmetrics	0.044	0.896	0.08	0.545		0.069	0.857	0.106	0.504		0.300	0.444	0.342	0.401	
	(0.077)	(0.107)	(0.088)	[0.91]		(0.104)	(0.126)	(0.114)	[0.70]		(0.145)	(0.200)	(0.177)	[0.89]	
GARCH(1,1)	0.052	0.882	0.087	0.546		0.083	0.787	0.165	0.508		0.300	0.349	0.436	0.430	
	(0.079)	(0.129)	(0.089)	[0.93]		(0.112)	(0.142)	(0.082)	[1.88]		(0.138)	(0.168)	(0.118)	[1.92]	
GARCH-D(1,1)	0.052	0.899	0.072	0.545		0.079	0.810	0.148	0.507		0.280	0.392	0.417	0.426	
	(0.08)	(0.132)	(0.092)	[0.71]		(0.113)	(0.146)	(0.083)	[1.74]		(0.141)	(0.172)	(0.118)	[2.01]	
GJR-G-D(1,1)	0.07	0.823	0.139	0.549		0.088	0.789	0.170	0.511		0.244	0.494	0.367	0.423	
	(0.087)	(0.146)	(0.098)	[1.02]		(0.117)	(0.183)	(0.115)	[0.90]		(0.146)	(0.208)	(0.153)	[1.23]	

Notes: The table presents estimates of regressions of realized standard deviation for the S&P 500 on a constant and either five, ten or twenty-days ahead forecasts from different models. The regression is $s_{t+h|t+1} = b_0 + b_1 \widehat{s}_{t+h|t,(1)} \; [+ b_2 \widehat{s}_{t+h|t,(2)}] + u_t$, where $\widehat{s}_{t+h|t,(i)}$ is the forecast of h-days realized standard deviation with $h \in \{5, 10, 20\}$ from model i. The forecast evaluation period covers January 2, 1998-December 29, 2000 ($R = 756$). See Table 6 for further details.

Table 9: Value-at-Risk evaluation for the out-of-sample period January 1998-December 2000

	MRC-μ	MRC-σ	$\hat{\alpha}$	LR_{uc}	LR_{ind}	LR_{cc}	C	$\overline{C}$	$\max(C_t)$
Riskmetrics	26.128	7.273	2.513	12.315	0.981	13.296	55.569	2.925	16.837
				(0.000)	(0.322)	(0.001)			
GARCH(1,1)	25.893	6.715	2.381	10.496	0.879	11.376	53.822	2.990	13.608
				(0.001)	(0.348)	(0.003)			
GARCH-D(1,1)	25.529	6.456	2.381	10.496	0.879	11.376	53.891	2.994	12.967
				(0.001)	(0.348)	(0.003)			
GJR-G-D(1,1)	24.113	5.773	2.249	8.791	0.783	9.574	46.883	2.758	8.740
				(0.003)	(0.376)	(0.008)			
ARFI	24.646	4.498	1.984	5.750	0.608	6.358	41.627	2.775	11.871
				(0.016)	(0.435)	(0.042)			
ARFI-D	24.673	4.531	2.249	8.791	0.783	9.574	45.087	2.652	14.225
				(0.003)	(0.376)	(0.008)			
ARFI-DR	24.646	4.507	2.381	10.496	0.879	11.376	51.123	2.840	18.516
				(0.001)	(0.348)	(0.003)			
ARFI-DRL	24.855	4.759	2.513	12.315	0.981	13.296	53.002	2.790	20.693
				(0.000)	(0.322)	(0.001)			
ARFI-DXR	25.008	4.789	2.381	10.496	0.879	11.376	50.397	2.800	17.695
				(0.001)	(0.348)	(0.003)			
ARFI-DXRL	25.563	5.201	2.381	10.496	0.879	11.376	46.216	2.568	16.897
				(0.001)	(0.348)	(0.003)			

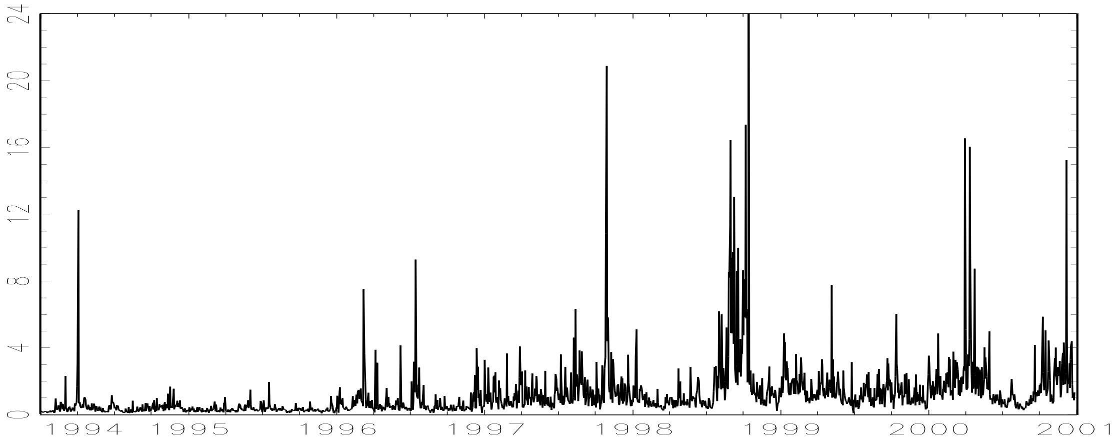
Notes: The table presents evaluation results for VaR estimates generated under the conditions of the Basle Committee MRA for the forecast evaluation period January 2, 1998-December 29, 2000 ($R = 756$). The first two columns show the average and standard deviation of the required capital to cover market risk exposure (in percentage terms). Column 3 shows the average percentage number of exceptions defined as $\hat{\alpha} = 100x/R$ where x is the number of exceptions. Columns 4-6 show the interval forecast evaluation test statistics of correct unconditional coverage (uc), independence (ind) and correct conditional coverage (cc) (p-values are between brackets). Columns 7-9 give the total score C based on (25), the average score, $\overline{C} = C/x$ and the maximum individual score (all in percentage terms).

Table 10: In-Sample Fit and Out-of-Sample Forecast Evaluation for daily DM/\$, ¥/\$ and ¥/DM realized volatility. The in-sample period is December 1, 1986-November 30, 1996 (2449 observations). The out-of-sample period is December 1, 1996-June 30, 1999 (596 observations)

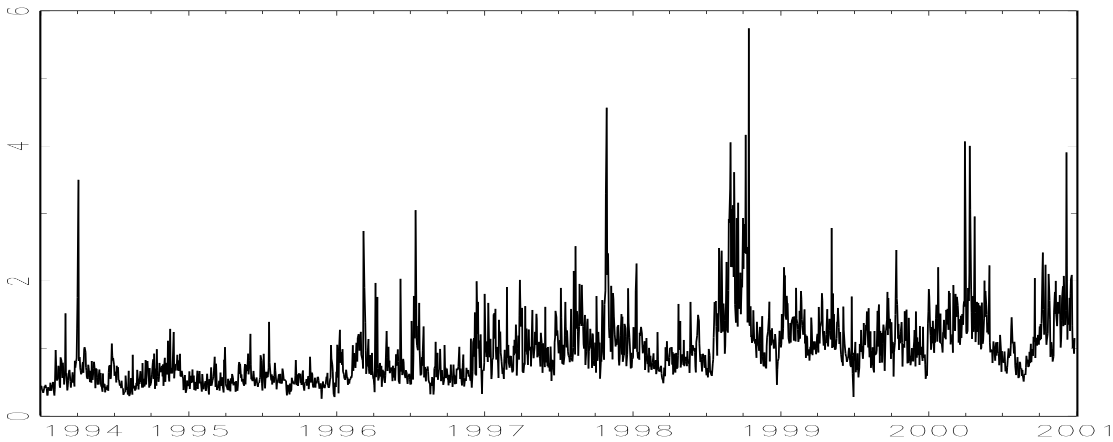
	In-Sample Fit		Out-of-Sample Forecast Evaluation							
			1-day		5-days		10-days		20-days	
	AIC	BIC	R^2	MSPE	R^2	MSPE	R^2	MSPE	R^2	MSPE
DM/\$										
Riskmetrics			0.107	0.030	0.084	0.024	0.054	0.022	0.044	0.021
GARCH(1,1)			0.101	0.027	0.078	0.018	0.048	0.015	0.036	0.012
GARCH-D(1,1)			0.094	0.027	0.076	0.018	0.047	0.015	0.034	0.012
GJR-G-D(1,1)			0.096	0.027	0.077	0.018	0.048	0.016	0.036	0.013
ARFI	-1.405	-1.394	0.326	0.020	0.278	0.014	0.229	0.012	0.158	0.011
ARFI-D	-1.482	-1.461	0.365	0.019	0.289	0.013	0.239	0.012	0.167	0.010
ARFI-DR	-1.504	-1.480	0.380	0.018	0.296	0.013	0.250	0.012	0.175	0.010
ARFI-DXR	-1.498	-1.474	0.379	0.018	0.290	0.013	0.230	0.011	0.150	0.010
ARFI-DRL	-1.502	-1.474	0.380	0.018	0.296	0.013	0.250	0.012	0.175	0.011
ARFI-DXRL	-1.500	-1.472	0.376	0.018	0.289	0.013	0.232	0.011	0.154	0.010
STARFI-DR	-1.504	-1.480	0.381	0.018	0.298	0.014	0.249	0.012	0.173	0.011
¥/\$										
Riskmetrics			0.238	0.116	0.235	0.091	0.224	0.084	0.245	0.073
GARCH(1,1)			0.256	0.122	0.240	0.099	0.221	0.093	0.249	0.086
GARCH-D(1,1)			0.245	0.124	0.225	0.102	0.208	0.096	0.233	0.089
GJR-G-D(1,1)			0.243	0.125	0.224	0.102	0.206	0.096	0.232	0.089
ARFI	-1.353	-1.345	0.366	0.093	0.319	0.073	0.272	0.067	0.254	0.058
ARFI-D	-1.404	-1.388	0.403	0.088	0.324	0.072	0.274	0.066	0.255	0.058
ARFI-DR	-1.437	-1.418	0.465	0.079	0.331	0.071	0.274	0.066	0.252	0.057
ARFI-DXR	-1.438	-1.419	0.490	0.078	0.347	0.074	0.286	0.073	0.260	0.072
ARFI-DRL	-1.439	-1.415	0.503	0.073	0.342	0.070	0.280	0.065	0.252	0.057
ARFI-DXRL	-1.447	-1.423	0.555	0.066	0.370	0.069	0.297	0.067	0.265	0.063
STARFI-DXRL	-1.449	-1.420	0.554	0.067	0.367	0.068	0.297	0.065	0.266	0.059
¥/DM										
Riskmetrics			0.241	0.099	0.268	0.081	0.285	0.075	0.341	0.068
GARCH(1,1)			0.245	0.101	0.261	0.084	0.272	0.079	0.333	0.073
GARCH-D(1,1)			0.243	0.101	0.259	0.084	0.270	0.079	0.329	0.072
GJR-G-D(1,1)			0.247	0.101	0.260	0.085	0.269	0.080	0.327	0.074
ARFI	-1.875	-1.863	0.351	0.068	0.302	0.053	0.269	0.048	0.269	0.041
ARFI-D	-1.954	-1.933	0.393	0.064	0.309	0.053	0.274	0.047	0.269	0.041
ARFI-DR	-1.971	-1.947	0.433	0.060	0.312	0.052	0.272	0.047	0.261	0.041
ARFI-DRL	-1.971	-1.942	0.469	0.057	0.320	0.051	0.276	0.047	0.266	0.041
ARFI-DXR	-1.970	-1.946	0.452	0.059	0.326	0.053	0.285	0.050	0.277	0.046
ARFI-DXRL	-1.975	-1.946	0.518	0.052	0.343	0.050	0.290	0.047	0.272	0.041
STARFI-DXRL	-1.980	-1.951	0.513	0.053	0.341	0.050	0.288	0.046	0.269	0.039

Notes: The table presents the in-sample fit and out-of-sample forecast evaluation for daily DM/\$, ¥/\$ and ¥/DM realized volatility. The first column shows the different models that were used for estimation and forecasting. The GARCH type models were only used for forecasting. The second and third column give the Akaike (AIC) and Schwarz (BIC) Information Criteria. Columns 4-11 show the evaluation of 1, 5, 10 and 20-day ahead forecasts of the realized standard deviation by means of the R^2 in the regression of the true realized standard deviation on a constant and the h-day ahead forecast for each of the models, and by means of the mean squared prediction error (MSPE).

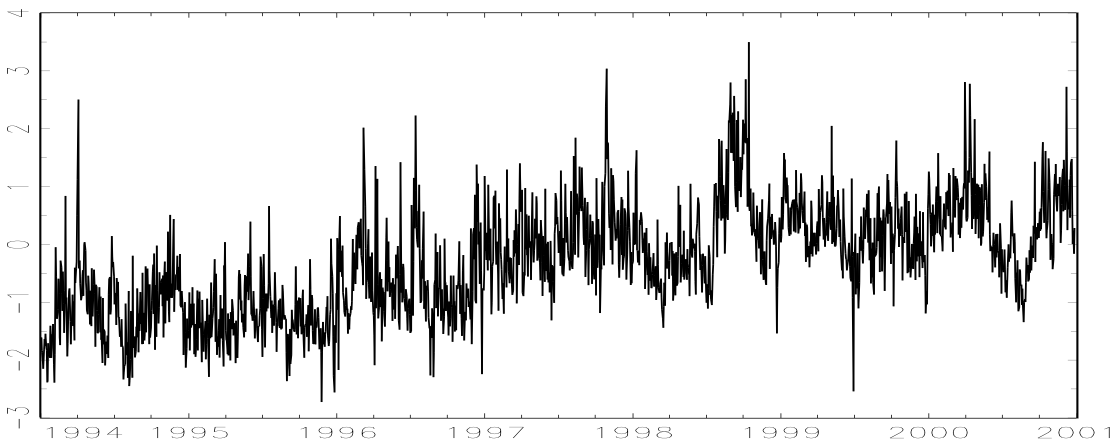
Figure 1: Realized S&P Volatility



(a) Realized variance



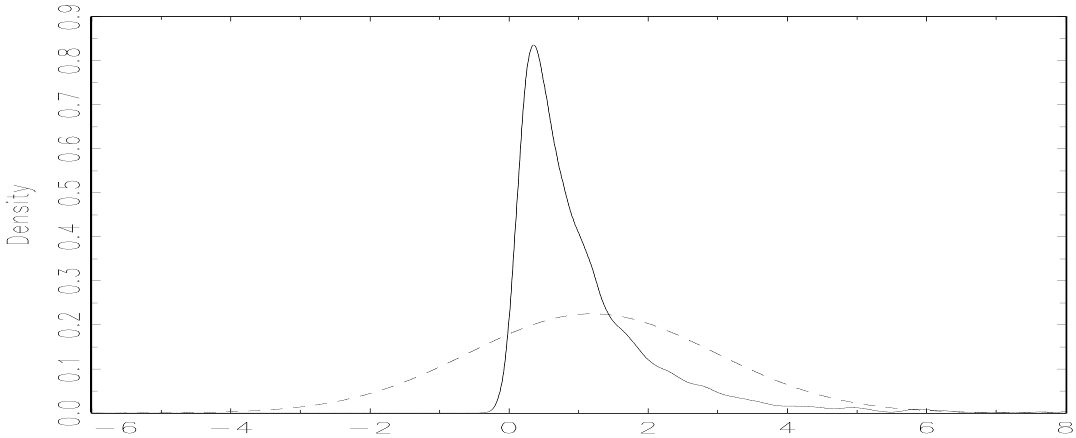
(b) Realized standard deviation



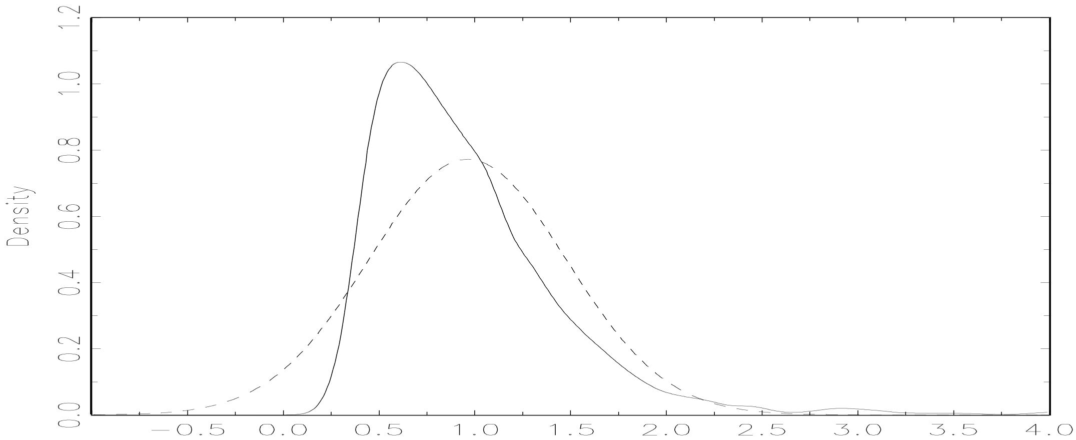
(c) Log realized variance

Notes: Daily realized volatility measures for S&P 500 returns, for the period from January 3, 1994, until December 29, 2000 (1767 observations).

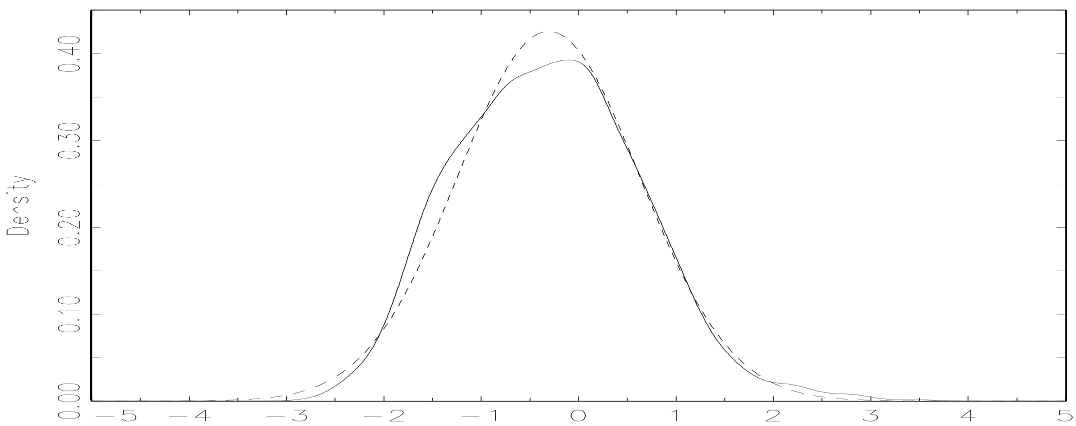
Figure 2: Realized Volatility Distributions



(a) Realized variance



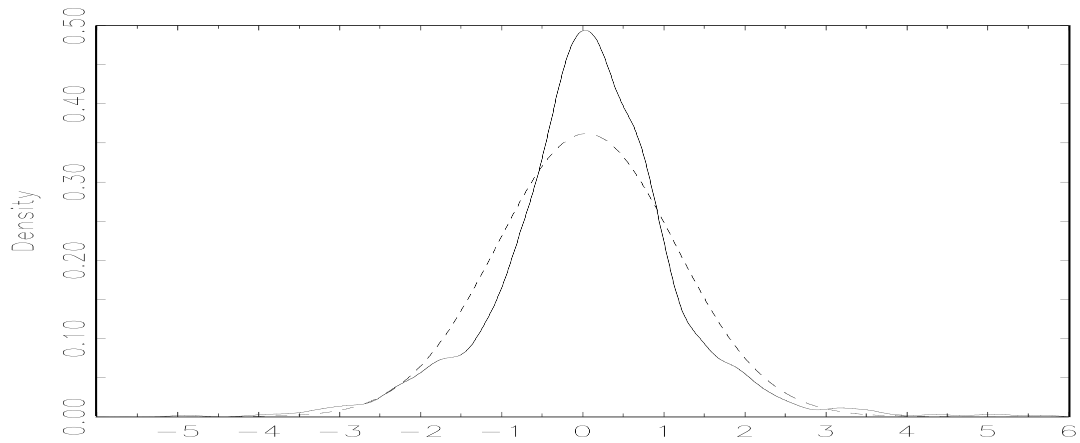
(b) Realized standard deviation



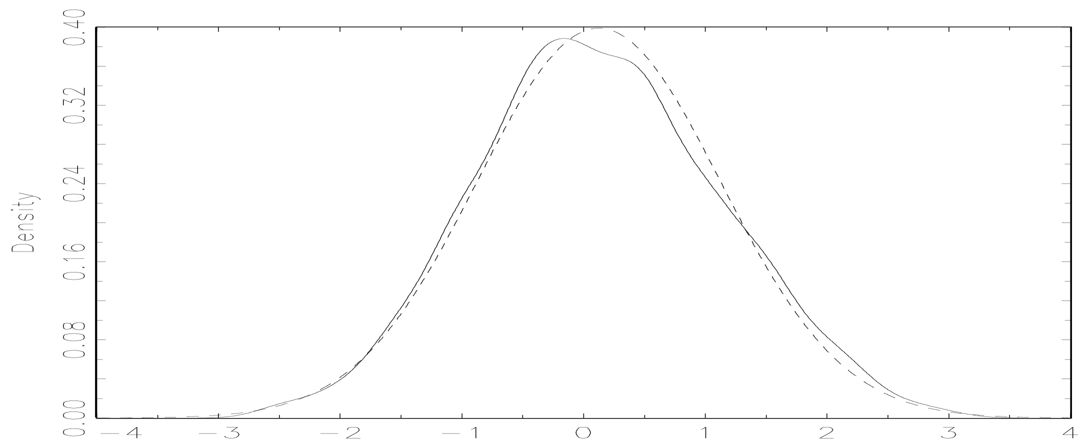
(c) Log realized variance

Notes: Kernel density estimates of daily realized volatility measures for S&P 500 returns, based on observations for the period from January 3, 1994, until December 29, 2000 (1767 observations). The solid line is the estimated density of the realized volatility measure (standardized to have zero mean and unit variance). The dashed line is a standard normal density.

Figure 3: Return Distributions



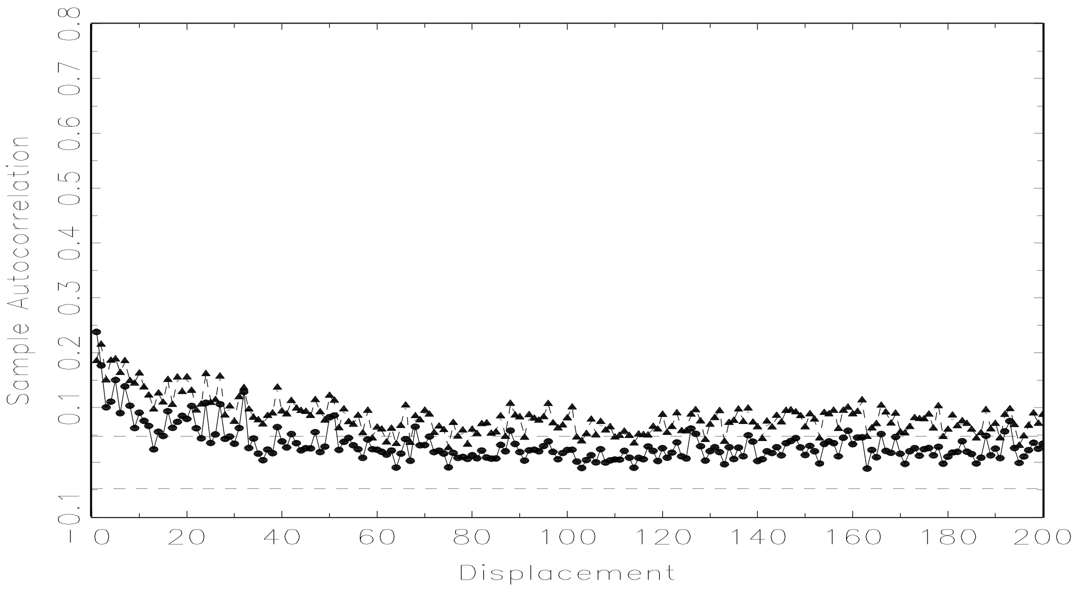
(a) Returns



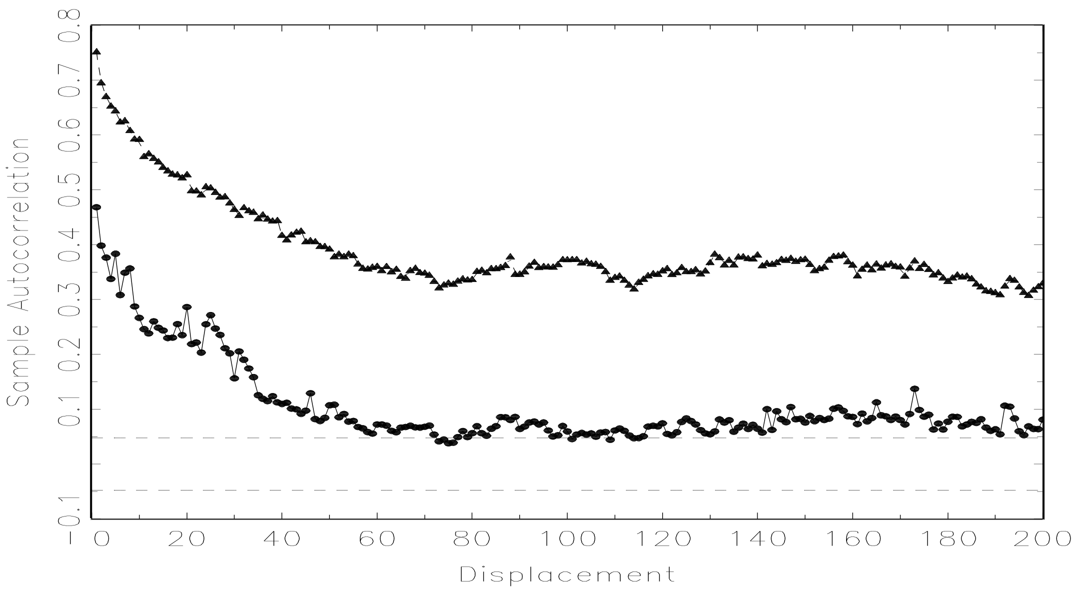
(b) Standardized returns

Notes: Kernel density estimates of daily S&P 500 returns r_t and standardized returns r_t/s_t (normalized to have zero mean and unit variance), based on observations for the period from January 3, 1994, until December 29, 2000 (1767 observations). The solid line is the estimated density of the returns (standardized to have zero mean and unit variance). The dashed line is a standard normal density.

Figure 4: Autocorrelation Functions for Returns and Realized Volatilities



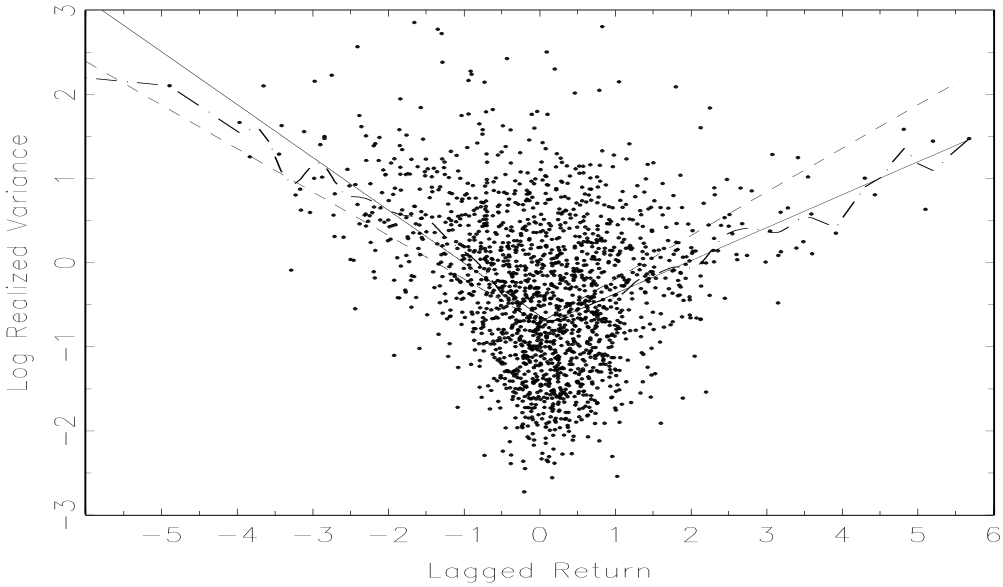
(a) Returns



(b) Realized volatilities

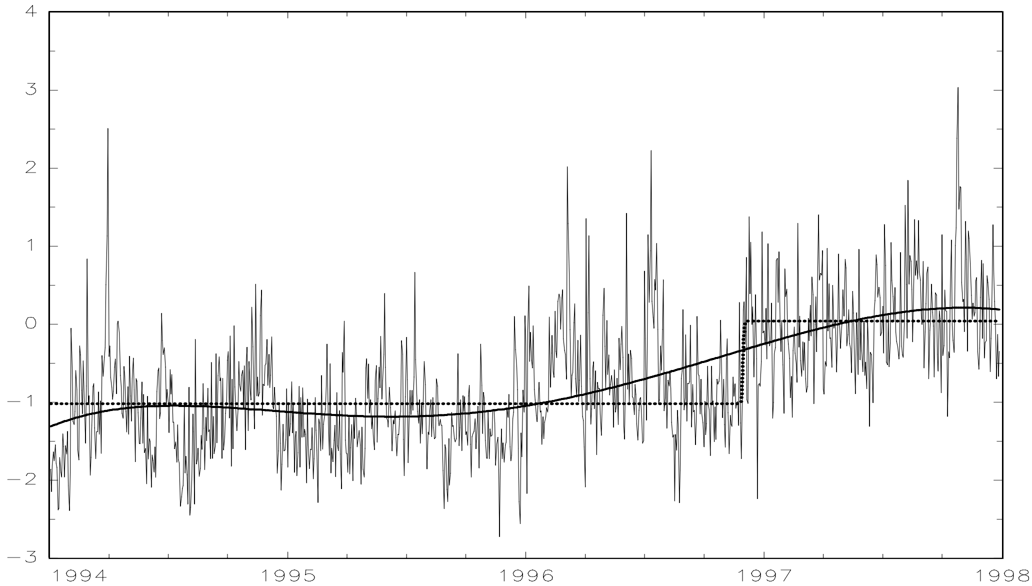
Notes: Panel (a) shows autocorrelation functions of daily squared returns (circles) and absolute returns (triangles), based on observations for the period from January 3, 1994, until December 29, 2000 (1767 observations). Panel (a) shows autocorrelation functions of daily realized variance (circles), realized standard deviation (triangles) and log realized variance (diamonds) for the same sample period. The dashed lines are Bartlett 95% confidence bounds, computed as $\pm 2/\sqrt{T}$, where T denotes the number of observations.

Figure 5: Leverage Effects in Realized Volatility



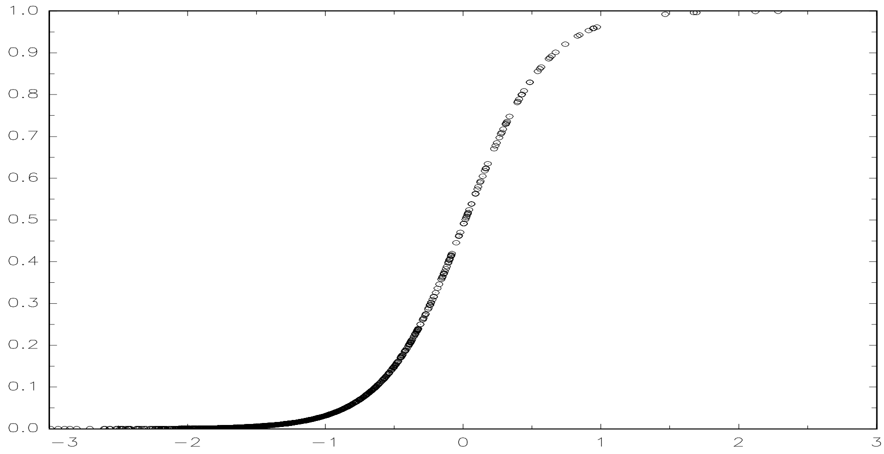
Notes: Scatter plot of daily log realized variance and lagged returns and realized volatility measures, based on observations for the period from January 3, 1994, until December 29, 2000 (1767 observations). The solid line is the fit of the news impact curve (2), where log realized volatility is regressed on a constant, the lagged absolute return, a dummy for negative returns and an interaction term of this dummy with the lagged absolute return. The dashed line is the fit of a symmetric news impact curve, i.e. (2) with $\beta_2 = \beta_3 = 0$. The dot-dashed line is the fit from a nonparametric regression of log realized volatility on the lagged return.

Figure 6: Modeling the structural break

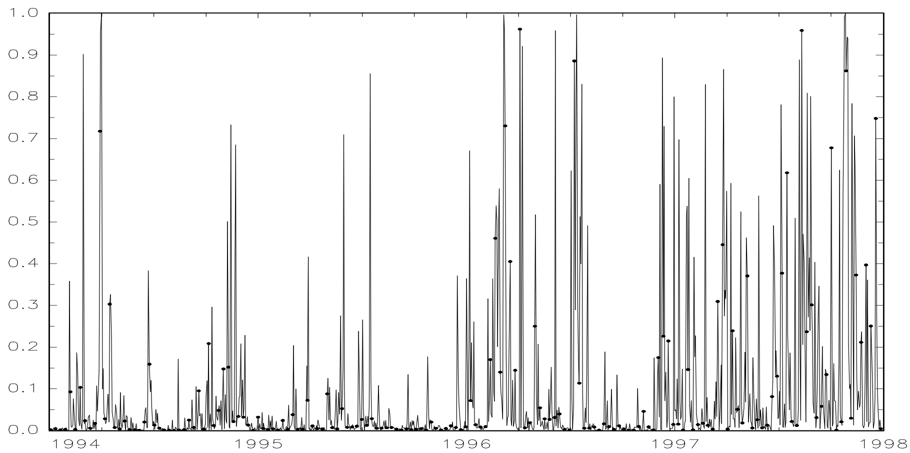


Notes: Plot of the daily log realized variance and the fit in the ARFI-SC model of $\alpha_0 + P(t)$ (solid line) and $\alpha_0 + \alpha_1 G(t; \gamma, \tau)$ (dotted line) with $P(t)$ and $G(t; \gamma, \tau)$ being the polynomial and logistic function respectively, for the in-sample period from January 3, 1994 until December 31, 1997.

Figure 7: Transition function



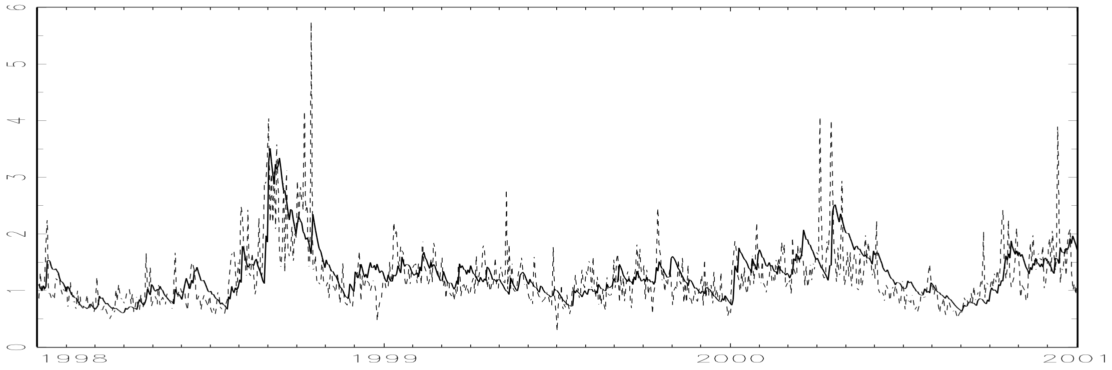
(a) Transition function versus the transition variable s_t



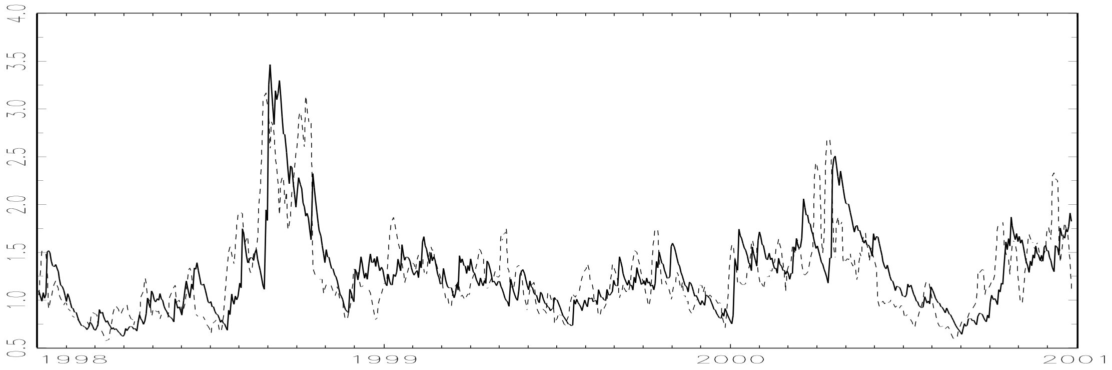
(b) Transition function over time

Notes: Transition function in STARFI-SCDXRL model against the transition variable and over time. The model was estimated for the in-sample period from January 3, 1994 until December 31, 1997.

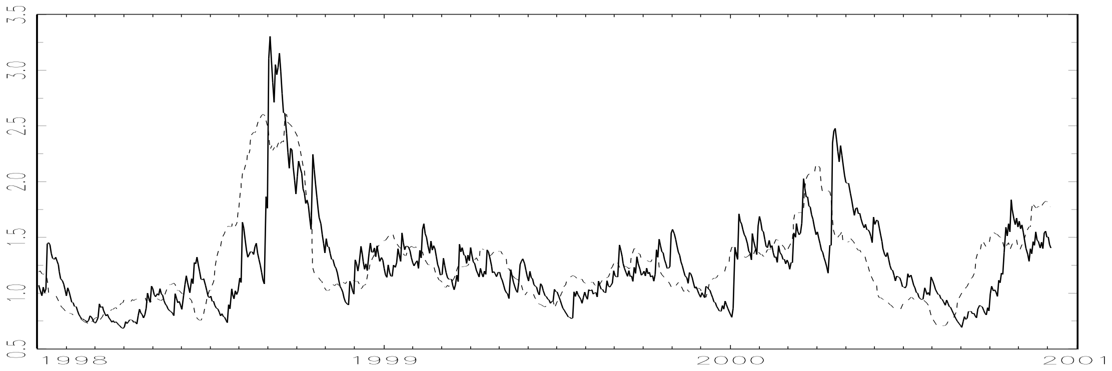
Figure 8: GARCH(1,1) Volatility Forecasts



(a) 1-day



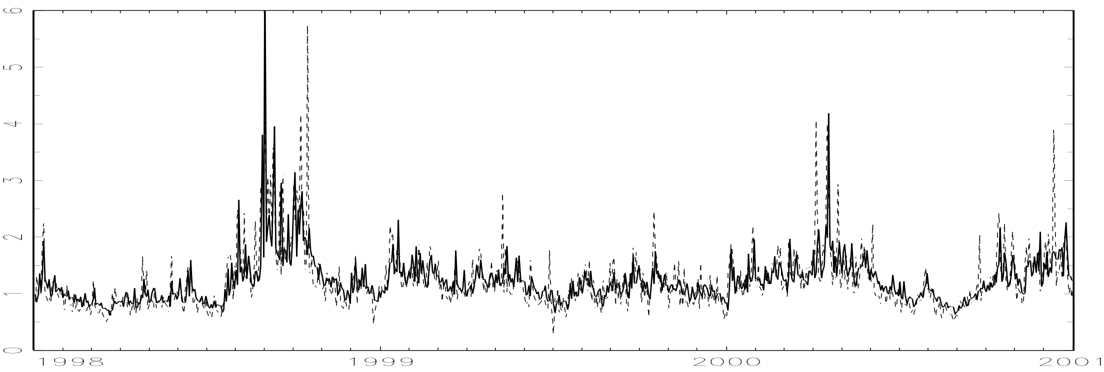
(b) 5-days



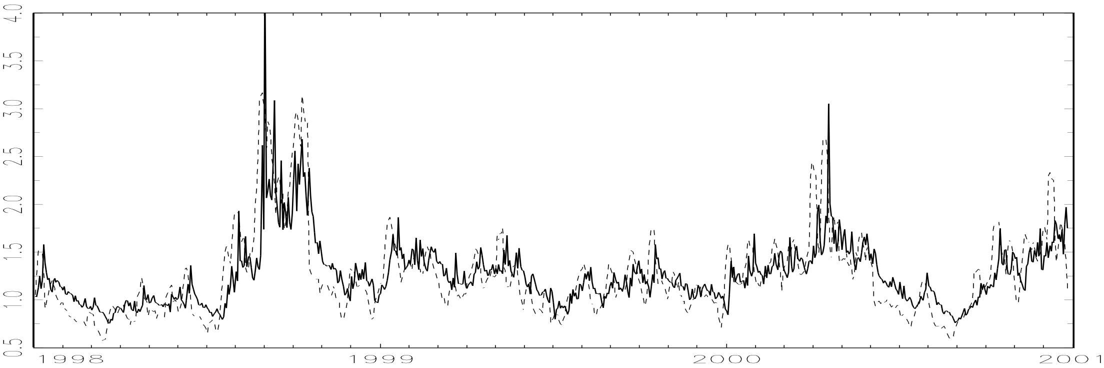
(c) 20-days

Notes: Out-of-sample forecasts $\widehat{s}_{t+h|t}$ of 1-day, 5-day and 20-day standard deviation obtained from a GARCH(1,1) model (solid lines) and realizations $s_{t+h|t+1}$ (dashed lines) for the period from January 2, 1998, until December 29, 2000 (756 observations for one-day ahead forecasts).

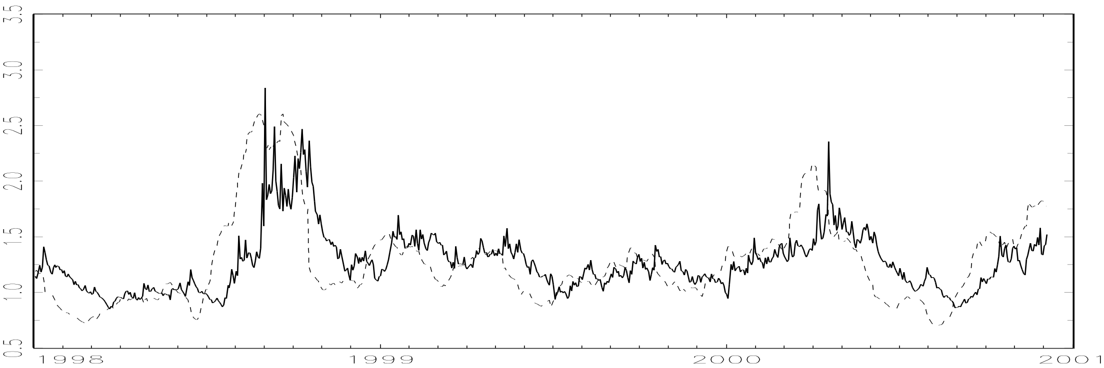
Figure 9: ARFI-DXRL Realized Volatility Forecasts



(a) 1-day



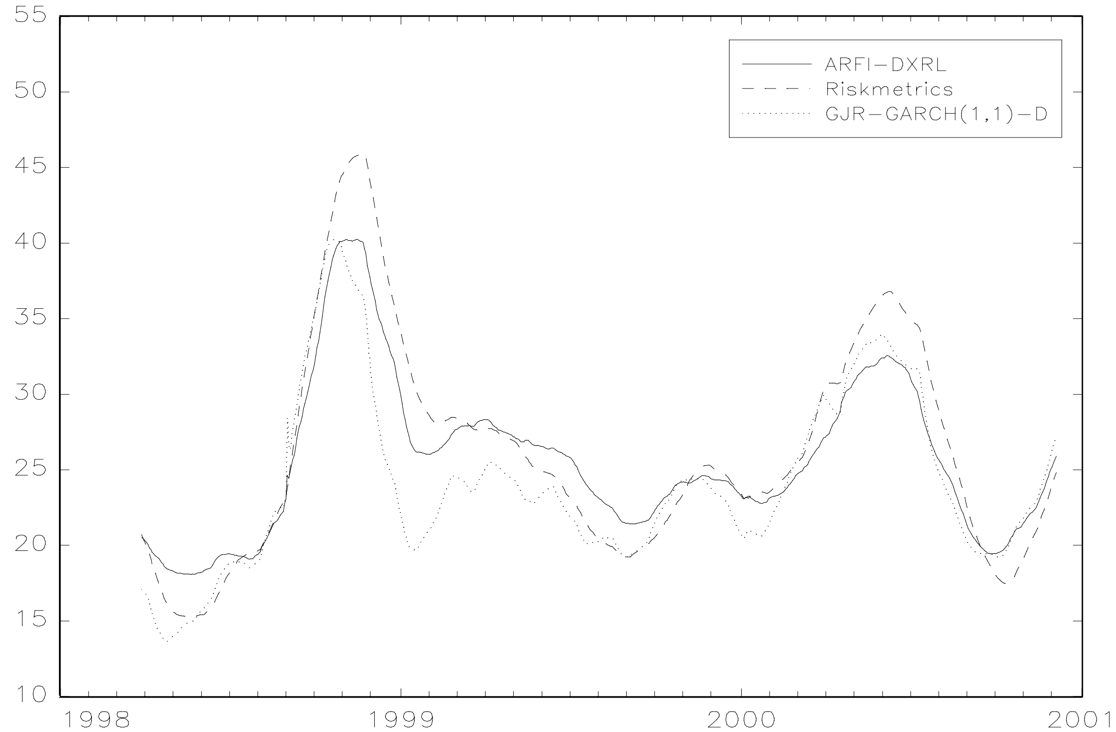
(b) 5-days



(c) 20-days

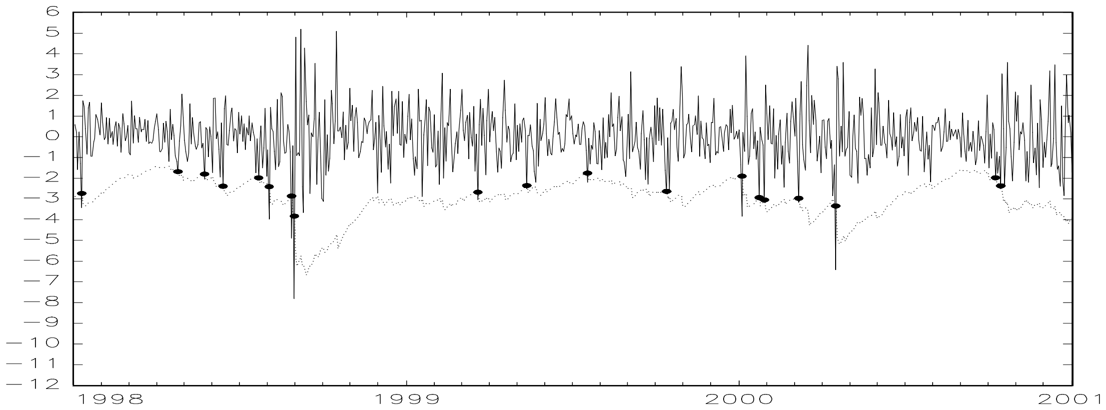
Notes: Out-of-sample forecasts $\widehat{s}_{t+h|t}$ of 1-day, 5-day and 20-day standard deviation obtained from the ARFI-DXRL model (solid lines) and realizations $s_{t+h|t+1}$ (dashed lines) for the period from January 2, 1998, until December 29, 2000 (756 observations for one-day ahead forecasts).

Figure 10: Market Risk Capital

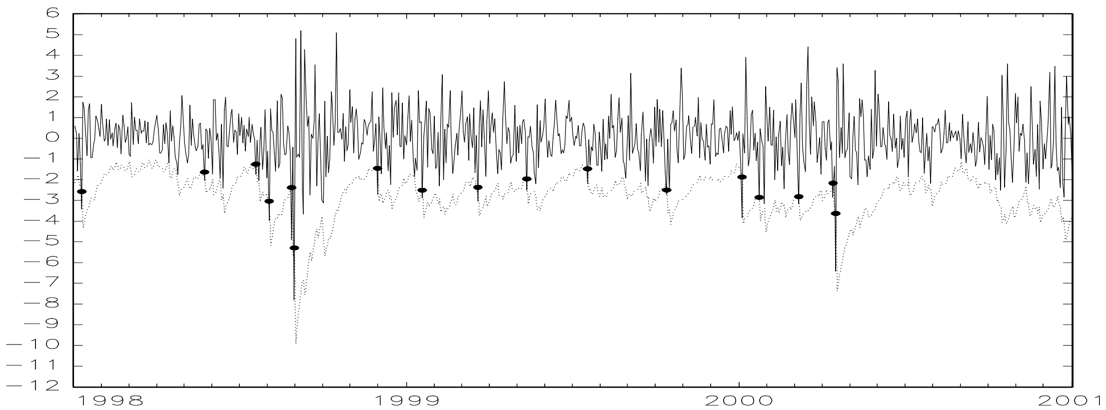


Notes: The graph shows the required capital (in percentage terms) to cover market risk exposure which is calculated as $MRC_{t+1} = \max(VaR_t^{10}, S_t \times \frac{1}{60}\sum_{i=0}^{59} VaR_{t-i}^{10})$ based on volatility forecasts from the Riskmetrics , GJR-GARCH(1,1)-D and ARFI-DXRL models from January 2, 1998, until December 29, 2000 (756 observations). VaR_t^{10} is the 99% VaR estimate for 10-day holding period. The first sixty 1-day VaR estimates were used to construct the initial history needed to calculate MRC_{t+1} which is the reason why the series starts at observation 60 of the out-of-sample period.

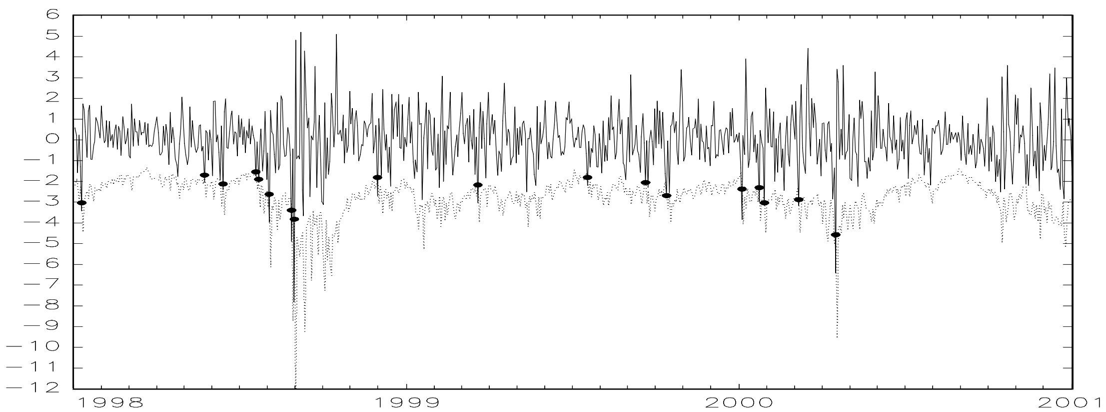
Figure 11: Value-at-Risk estimates



(a) Riskmetrics



(b) GJR-GARCH(1,1)-D



(c) ARFI-DXRL

Notes: Realized returns (solid line) and 1-day 99% Value-at-Risk estimates based on volatility forecasts for the Riskmetrics, GJR-GARCH(1,1)-D and ARFI-DXRL models (dotted lines) for the period from January 2, 1998, until December 29, 2000 (756 observations). Black dots indicate the VaR exceptions.