Title

Exploratory Analysis and Modeling of Stock Returns

Journal

Journal of Computational and Graphical Statistics, 25(2)

ISSN

1061-8600

Authors

Noguchi, Kimihiro
Aue, Alexander
Burman, Prabir

Publication Date

2016-04-02

DOI

10.1080/10618600.2014.988338

Exploratory Analysis and Modeling of Stock Returns

Kimihiro Noguchi Alexander Aue Prabir Burman*

Abstract

In this paper, novel joint semiparametric spline-based modeling of conditional mean and volatility of financial time series is proposed and evaluated on daily stock return data. The modeling includes functions of lagged response variables and time as predictors. The latter can be viewed as a proxy for omitted economic variables contributing to the underlying dynamics. The conditional mean model is additive. The conditional volatility model is multiplicative and linearized with a logarithmic transformation. In addition, a cube-root power transformation is employed in order to symmetrize the lagged response variables. Using cubic splines, the model can be written as a multiple linear regression, thereby allowing predictions to be obtained in a simple manner. As outliers are often present in financial data, reliable estimation of the model parameters is achieved by trimmed least squares (TLS) estimation for which a reasonable amount of trimming is suggested. To obtain a parsimonious specification of the model, a new model selection criterion corresponding to TLS is derived. Moreover, the (three-parameter) generalized gamma distribution is identified as suitable for the absolute multiplicative errors and shown to work well for predictions and also for the calculation of quantiles, which is important to determine the value at risk. All model choices are motivated by a detailed analysis of IBM, HP, and SAP daily returns. The prediction performance is compared to the classical GARCH and APGARCH models as well as to a nonstationary time-trend volatility model. The results suggest that the proposed model may possess a high predictive power for future conditional volatility.

*K. Noguchi (E-mail: Kimihiro.Noguchi@wwu.edu) is Assistant Professor, Department of Mathematics, Western Washington University, Bellingham, WA 98225. Alexander Aue (E-mail: aaue@ucdavis.edu) is Associate Professor and Prabir Burman (E-mail: pburman@ucdavis.edu) is Professor, Department of Statistics, University of California, One Shields Avenue, Davis, CA 95616. This research was partially supported by NSF grants DMS 09-05400, DMS 09-07622, DMS 12-09226, DMS 13-05858, and DMS 14-07530.

KEYWORDS: Conditional mean, Conditional volatility, Financial time series, Prediction

1 Introduction

Modeling the square root of the conditional expected squared fluctuation from the conditional mean in stock returns, commonly referred to as conditional volatility, plays an important role in risk management. To overcome the unrealistic assumption of constant one-step-ahead conditional volatility, Engle (1982) proposed the autoregressive conditional heteroscedastic (ARCH) model which relates the current squared conditional volatility to the past squared fluctuations from the conditional mean. Further extensions of the class of ARCH models were for example suggested by Bollerslev (1986), Nelson (1991), and Glosten, Jagannathan, and Runkle (1993), a summary of which can be found in Aue, Berkes, and Horváth (2006).

These classical financial time series models are constructed on the assumption of stationarity of the financial time series of interest. However, empirical evidence often suggests the presence of nonstationarity especially if the length of the collected data is relatively large (see Mikosch and Stărică (2004), Fryzlewicz, Sapatinas and Subba Rao (2006), and references therein). To capture such nonstationarity, recent contributions to the literature (Dahlhaus and Subba Rao (2006); Engle and Rangel (2008); Vogt (2012)) proposed the inclusion of a smooth function of (rescaled) time into the model, along with other predictors such as smooth functions of lagged response variables in the context of conditional volatility modeling. As a consequence, the response variable is assumed to behave in a locally stationary manner, and the function of time can be interpreted as a proxy for economic variables not considered in the model.

In this paper, a novel robust semiparametric spline-based approach is proposed for the joint modeling of conditional mean and volatility of the form

$$r_t = \mu_t + \sigma_t \varepsilon_t,$$

where r_t, μ_t, σ_t, and ε_t denote return, conditional mean, conditional volatility, and multiplicative error, respectively. The modeling approach aims at achieving com-

putationally inexpensive and reliable predictions of short-run volatility and at the same time capturing the tail behavior of the fluctuations from the conditional mean accurately. Firstly, both conditional mean and volatility models include a smooth function of time as a predictor in addition to functions of lagged response variables. The implication is that both μ_t and σ_t may behave in a nonstationary manner. Each function is modeled in a nonparametric manner by a cubic spline, which admits a linear regression form, to make predictions computationally feasible. Secondly, similar to Yang, Härdle, and Nielsen (1999), an additive structure is imposed on the conditional mean and a multiplicative structure on the conditional volatility. Akin to Nelson's (1991) EGARCH, the conditional volatility model is linearized by taking the natural logarithm, and thus the predictions are made based on log conditional volatility. Thirdly, to capture the tail behavior of the fluctuations accurately, $|\varepsilon_t|$ is modeled by the (three-parameter) generalized gamma distribution, noting that a cube-root transformation provides reasonable symmetry.

To achieve reliable predictions in the presence of outliers often contaminating financial time series data, symmetry of the response variables and trimming of outliers are considered. First, coefficients of the cubic splines are estimated by trimmed least squares (TLS). A reasonable amount of trimming is suggested and a new Mallows' C_p-type model selection criterion corresponding to TLS estimates is developed to achieve parsimony. Second, appropriate power transformations are suggested to approximately symmetrize the distributions of lagged response variables. In particular, a cube-root transformation is suggested for the absolute fluctuations and absolute multiplicative error. These ideas are backed by extensive empirical observations of financial time series from various industry sectors, and the proposed modeling approach seems to work well across a variety of financial time series encountered in practice.

The proposed modeling approach borrows parts of ideas from several previously established approaches. The additive/multiplicative structure is similar to that of Yang, Härdle, and Nielsen (1999) and Kim and Linton (2004) who generalized it to a generalized additive structure, assuming stationarity of the process of interest.

Their models are based on an application of local polynomial regression with marginal integration (see Linton and Nielsen (1995) and references therein). Moreover, local polynomial-based conditional volatility models assuming locally stationary processes were proposed, for example, by Dahlhaus and Subba Rao (2006), and Vogt (2012). Their approaches typically require a number of iterative procedures for making short-run predictions, and hence are computationally expensive.

The proposed conditional volatility model is compared to GARCH-type models and to a model similar to the one proposed in Fryzlewicz, Sapatinas and Subba Rao (2006) in terms of its predictive performance using daily stock returns of IBM, HP, and SAP traded at the New York Stock Exchange (NYSE). The observations span 1100 days (and hence 1099 returns) from January 16, 2004 to May 30, 2008, and were obtained from Yahoo! Finance.

The remainder of this paper is organized as follows. The modeling strategy and analysis of the conditional mean and conditional volatility are described in Sections 2 and Section 3. The predictive performance of the proposed model is evaluated and compared to competitor approaches in Section 4. Some technical aspects of the proposed prediction methodology are given in Section 5. Conclusions are offered in Section 6.

2 Conditional mean modeling

2.1 Summary statistics for daily returns

An important yet sometimes ignored feature of financial time series analysis is the conditional mean modeling. The conditional mean of r_t is given by $\mu_t = E[r_t|\mathcal{F}_{t-1}]$, so that

$$r_t = \mu_t + a_t,$$

where $\mathcal{F}_{t-1}$ is the information set available up to time $t-1$ and a_t the residual from μ_t. The time series plots of the IBM, HP, and SAP return series given in Figure 1 indicate that each r_t has approximately mean zero. All the plots show a number of outliers.

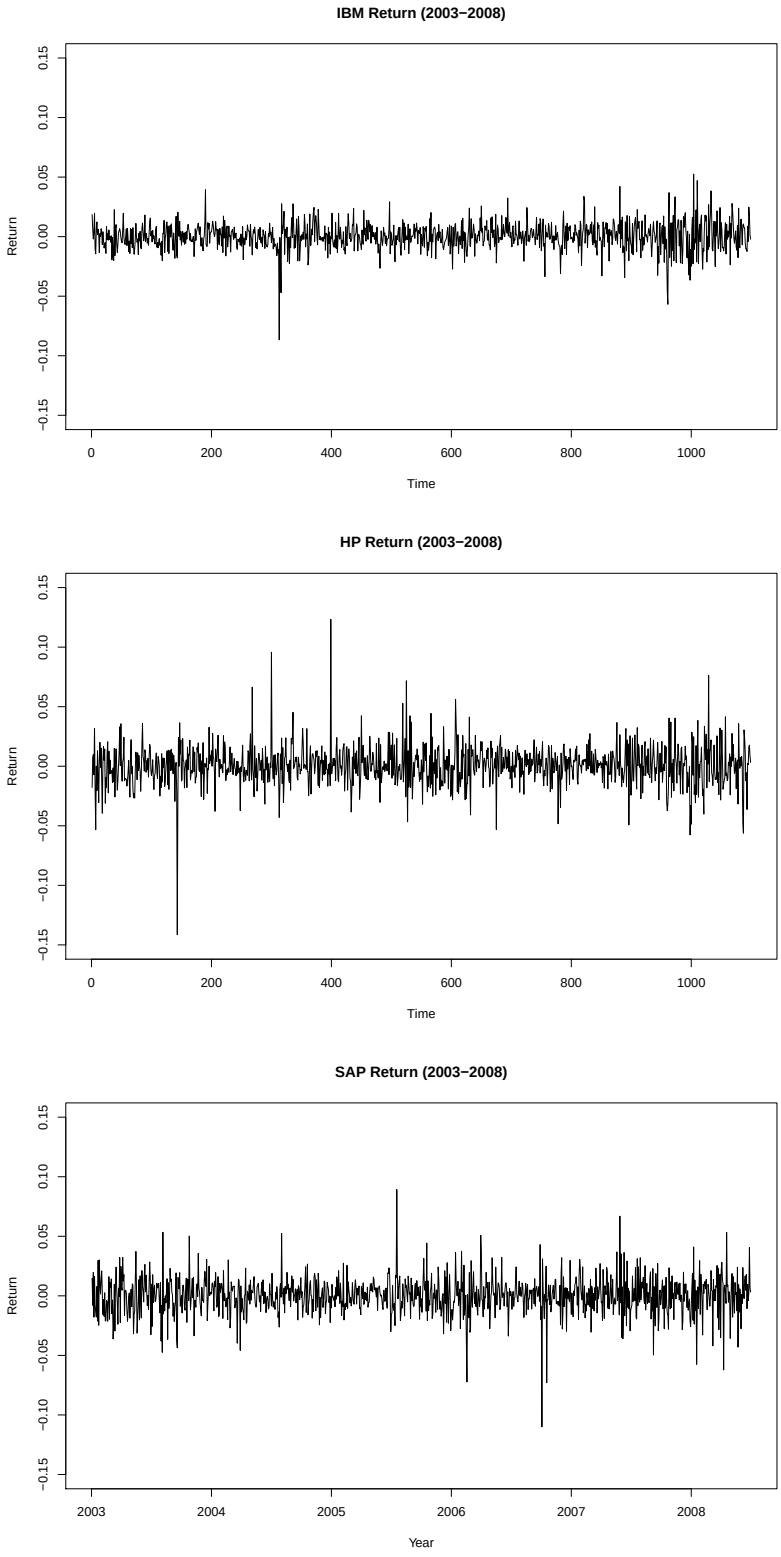

Figure 1: Time series plots of three stock returns.

Statistic	**Skewness**		**Excess Kurtosis**	
Data	**Original**	**Trimmed**	**Original**	**Trimmed**
IBM	-0.58	0.00	5.98	0.09
HP	-0.20	0.03	10.15	0.23
SAP	0.44	0.06	7.92	0.12

Table 1: Skewness and excess kurtosis of r_t for the original and the lower and upper 1% trimmed data.

While the majority of the series tends to look stationary, there also seem to be periods where the mean of the returns appears to have some time-varying movement. This suggests that the inclusion of a smooth function of time may be beneficial to improve the overall performance of conditional mean and volatility predictions.

Table 1 shows a pattern consistent with previous findings of Franke, Härdle, and Hafner (2008): The return r_t in general has skewness very close to zero and excess kurtosis greater than zero. However, for trimmed data (both the lower and upper 1% have been deleted), the excess kurtosis becomes very close to zero as well, implying that high excess kurtosis in returns is mainly influenced by a number of extreme observations.

Assuming that r_t is a stationary linear process, the order of its linear dependence can be determined from the sample ACF, PACF, and some information criterion such as AIC (Akaike, 1974). While the stationarity assumption may not hold, for example, when a deterministic trend exists, sample ACF and PACF are used as initial guidance to identify orders of return autocorrelation with the lower and upper 1% of the data trimmed for robustness. Since the analysis suggests that the majority of significant lags is within this range, $p_1 = 10$ lags are chosen for the full model.

2.2 Conditional mean specification and estimation

Based on the data analysis reported in the previous section, a conditional mean model of the form

$$\mu_t = f_0(t/T) + f_1(r_{t-1}) + f_2(r_{t-2}) + \cdots + f_{p_1}(r_{t-p_1}), \quad t = p_1 + 1, \ldots, T,$$

is entertained. The functions $f_0, f_1, \ldots, f_{p_1}$ are modeled with cubic splines, i.e.,

$$f_i(x) = \beta_{i,0} + \beta_{i,1}x + \beta_{i,2}x^2 + \beta_{i,3}x^3 + \sum_{j=4}^{k_i+3} \beta_{i,j} \max\{(x - x_{i,j})^3, 0\},$$

where k_i knots are placed at $x_{i,4}, x_{i,5}, \ldots, x_{i,k_i+3}$ for $i = 0, 1, \ldots, p_1$. For identifiability, $\beta_{i,0} = 0$ for $i = 1, 2, \ldots, p_1$. This leads to the linear representation

$$\mathbf{r} = X\boldsymbol{\beta} + \mathbf{a},$$

where $\mathbf{r} = (r_{p_1+1}, \ldots, r_T)'$, $\mathbf{a} = (a_{p_1+1}, \ldots, a_T)'$, and $\boldsymbol{\mu} = X\boldsymbol{\beta}$ is the vector representation of the conditional mean. For the function of rescaled time f_0, two knots are placed at 1/3 and 2/3. For the functions of lagged response variables $f_1, \ldots, f_{p_1}$, three knots are placed at $r_{i,16}$, $r_{i,50}$, and $r_{i,84}$, where $r_{i,q}$ is defined to be the lower qth percentile of r_{t-i}. These percentiles are chosen to mimic knot placements at the sample mean and sample mean plus/minus standard deviation for normal variables, albeit in a robust way. In order to derive theoretical results, one places a dense grid of knots in the range of the independent variables (here time and lags), and knot selection becomes equivalent to model selection which can be performed by any available information criterion. For practical purposes, one can often (but not always) get away with placing only a few knots, in particular if there are no sharp changes in the variables. From visual inspections of the data, this seems to be the case here. Moreover, all prediction results to be reported below appear to be influenced only minorly by the knot placement.

Let m_r be the sample median of r_t, and define $\tilde{r}_t = r_t - m_r$. To mitigate the effect of outliers, the weights

$$w_{q,t} = \begin{cases} 1, & \text{if } |\tilde{r}_t| < c_q, \\ 0, & \text{otherwise,} \end{cases}$$

where c_q is the lower $(2q-100)$th percentile of $|\tilde{r}_t|$, are introduced. For example, for $q = 99$, c_{99} is the lower 98th percentile of $|\tilde{r}_t|$. The resulting weights are analogous to deleting $\tilde{r}_t$ outside the range between the 1st and 99th percentiles, assuming that the distribution of $\tilde{r}_t$ is symmetric around zero. To obtain robust estimates, let W_q denote a diagonal matrix with diagonal entries $(w_{q,p_1+1}, \ldots, w_{q,T})$. Then, assuming that $E[W_q\mathbf{a}] = \mathbf{0}$ and $E[\mathbf{a}'W_q\mathbf{a}] < \infty$, the TLS estimator of $\boldsymbol{\beta}$ is given by

$$\hat{\boldsymbol{\beta}} = (X'W_qX)^{-1}X'W_q\mathbf{r}. \tag{2.1}$$

Letting $\hat{\boldsymbol{\mu}} = (\hat{\mu}_{p_1+1}, \ldots, \hat{\mu}_T)'$, the conditional mean is estimated by $\hat{\boldsymbol{\mu}} = X\hat{\boldsymbol{\beta}}$. Some theoretical properties of this estimator are are established in Section 5.

Putting such weights is not equivalent to deleting extreme observations from the dataset because the weights are determined solely based on the response variables but not the lagged response variables for the predictors. In other words, it is possible, for example, that r_{t-1} for which $w_{q,t-1} = 0$ may serve as a predictor as long as r_t has the corresponding weight $w_{q,t} = 1$. This way the model is able to adapt to unusual predictor values and forecasting performance is improved.

2.3 Model selection

It is quite natural to suspect that many of the predictors used in the conditional mean formulation are correlated. This could cause instability in the parameter estimation. To obtain a more parsimonious expression, methods of nonparametric lag selection based on cross-validation (Gao and Tong, 2004) or nonparametric final prediction error type criteria (Tschernig and Yang, 1999) have been developed. However, as Rech, Teräsvirta, and Tschernig (2001) pointed out, these model selection techniques tend to be computationally expensive. Therefore, a model selection procedure based on a Mallows' C_p-type criterion for TLS is developed in this section which allows for reasonable predictions of μ_t without adding a high computational burden.

Burman and Nolan (1995) suggested a modification of Mallows' C_p criterion which is applicable to various types of loss functions. In the case of TLS, assuming that $\tilde{r}_t$

has a symmetric distribution around zero and $\hat{\boldsymbol{\beta}} \in \mathbb{R}^d$ (so that a subset of d predictors is chosen from the full model), the modified Mallows' C_p-type criterion

$$M_{W_q} = \left(1 + \frac{2d}{\sum_{t=p_1+1}^{T} w_{q,t}}\right) \sum_{t=p_1+1}^{T} (r_t - \hat{\mu}_t)^2 w_{q,t}$$

is obtained. Notice that the full model uses a fixed number of 10 lags, each of which is associated with six parameters. There are another five parameters associated with the function of time, so that the full model has 65 parameters to choose from. The following stepwise model selection procedure may be utilized.

1. Center r_t by subtracting its sample median m_r, yielding $\tilde{r}_t = r_t - m_r$.
2. Starting from the constant mean model, perform a forward selection repeatedly using M_{W_q} with $\tilde{r}_t$ as the response variable, in each step including that variable into the model which produces the greatest reduction in the M_{W_q} value. Stop if no more such reduction occurs.
3. Once the forward selection process is complete, perform a backward deletion repeatedly using M_{W_q}, in each step deleting that variable from the model which produces the greatest reduction in the M_{W_q} value. Stop if no more such reduction occurs.

Regarding the choice of q, recall that Table 1 suggests $q = 99$ to be reasonable as the effect of extreme observations appears to vanish by comparing the changes in excess kurtosis. This choice is further corroborated by visual inspection of the time series plots of μ_t (omitted for brevity) using various values of q. For this choice, Table 2 displays a summary of the number of parameters selected. In all three cases, time parameters are chosen (t/T for IBM, t/T and $(t/T)^2$ for HP, and $\max\{(t/T - 2/3)^3, 0\}$ for SAP), indicating a time-varying pattern in returns and the importance of including a function of time for relatively short financial time series data for an improved prediction performance. The application of M_{W_q} also leads to a model with a considerably smaller number of predictors compared to the full model.

Data	Time	Lags	Total
IBM	1	9	10
HP	2	12	14
SAP	1	8	9
Full Model	5	60	65

Table 2: Number of parameters selected using $q = 99$.

3 Conditional volatility modeling

3.1 Power transformations of $|a_t|$

Modeling and predicting conditional volatility σ_t are of central importance to financial time series analysts. An application is found in assessing the value-at-risk (VaR), an extreme quantile of r_t, which requires an accurate estimate of the conditional volatility at time t and the underlying distribution of a_t. A standard approach for modeling σ_t is to consider a multiplicative error structure

$$a_t = \sigma_t \varepsilon_t,$$

where ε_t are independent and identically distributed (i.i.d.) random variables with $E[\varepsilon_t] = 0$ and $E[\varepsilon_t^2] = 1$. Assuming that ε_t has a symmetric distribution, one can focus on the *magnitude* of a_t, namely,

$$|a_t|^\delta = \sigma_t^\delta |\varepsilon_t|^\delta,$$

for some $\delta > 0$. Since Engle (1982), the behavior of conditional volatility has typically been analyzed through a_t^2 (that is, $\delta = 2$). However, a_t^2 tends to be highly positively skewed, and hence, not an ideal choice for least squares estimation in a linear model. Another route is to use $|a_t|$ (that is, $\delta = 1$), as Ding, Granger and Engle (1993) find that $|a_t|$ often maximizes autocorrelation (see Penzer, Wang, and Yao (2009) and references therein). This is commonly known as the *Taylor effect*.

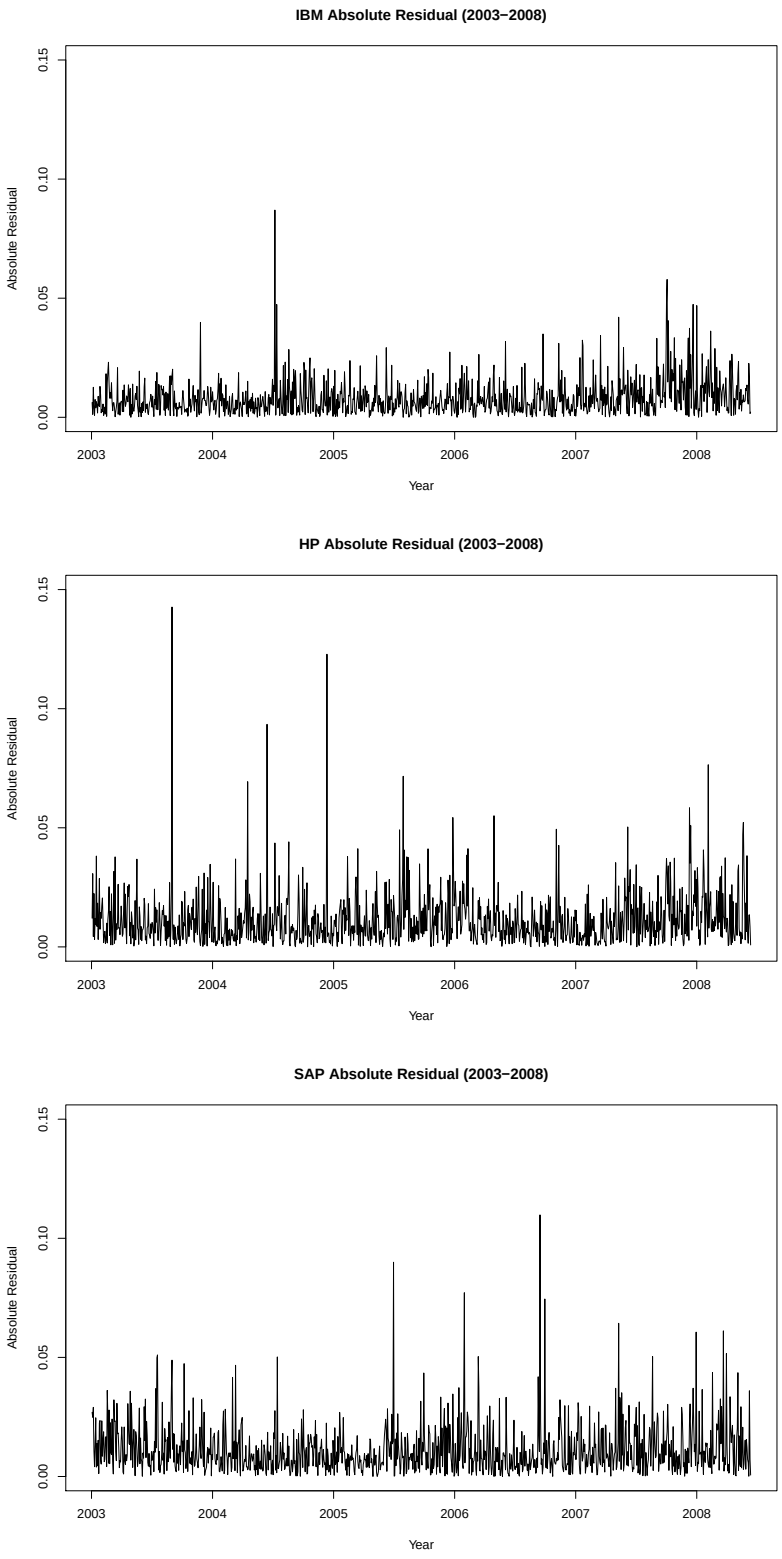

Figure 2: Time series plots of $|a_t|$ for three stocks.

Data	δ
IBM	0.31
HP	0.27
SAP	0.30

Table 3: The selected δ for which S_q is minimized for $|a_t|^\delta$.

For reliable parameter estimation, it is often beneficial to find a value of δ for which the underlying distribution of $|a_t|^\delta$ is symmetrized. This is the approach taken in this paper and is done here using the quantile-matching technique which selects the δ that minimizes the criterion function

$$S_q = \sum_{i \in U} \frac{|L_i - R_i|}{\text{IQR}}, \tag{3.1}$$

where L_i and R_i are the distances of the i-th and $(100-i)$-th percentile from the sample median of $|a_t|^\delta$, respectively. The denominator IQR stands for the interquartile range of $|a_t|^\delta$ which is used for standardization. To cover a broad range of percentiles, $U = \{1, 10, 20, 30, 40\}$ is used. The minimum of S_q is found by a grid search, varying δ from 0.01 to 2.00 with an increment of 0.01. It can be seen from Table 3 that, in all cases, $\delta \approx 1/3$ appears to minimize S_q well. This suggests a hitherto unobserved stylized fact of financial time series which may be coined the *symmetry effect*. Subsequently, the cube-root transformation $\delta = 1/3$ is applied in the model building process.

3.2 Conditional volatility specification and estimation

Let ε_t be symmetric i.i.d. random variables such that $E[\varepsilon_t] = 0$ and $E[|\varepsilon_t|^\delta] < \infty$. Utilizing the results of Section 3.1 and a log transformation, conditional volatility is modeled by

$$\begin{aligned} |a_t|^\delta &= \sigma_t^\delta |\varepsilon_t|^\delta, \\ \log \sigma_t &= h_0(t/T) + h_1(|a_{t-1}|^\delta) + \cdots + h_{p_2}(|a_{t-p_2}|^\delta), \end{aligned} \tag{3.2}$$

Data	Time	Lags	Total
IBM	1	4	5
HP	1	5	6
SAP	0	8	8
Full Model	5	60	65

Table 4: Number of parameters selected using $q = 99$.

where $\delta = 1/3$, $t = p_1 + p_2 + 1, \ldots, T$. Conducting a similar analysis as in Section 2.2 using volatilities detrended by local running medians to avoid an overshooting of the order due to potential nonstationarities, $p_2 = 10$ is chosen. The functions $h_0, h_1, \ldots, h_{p_2}$ are modeled with cubic splines following the same knot placements as in Section 2.2. For identifiability, similar restrictions as in the conditional mean model in Section 2.2 are imposed for h_i, $i = 1, \ldots, p_2$. Estimation is then performed by TLS using $q = 99$. A smooth function of time h_0 is included because Figure 2 shows some periods with time-varying volatility, commonly known as volatility clustering.

Predictors are included into the model with the selection procedure M_{W_q} developed in Section 2.3 based on the linear regression

$$\log|a_t| = h_0(t/T) + h_1(|a_{t-1}|^\delta) + \cdots + h_{p_2}(|a_{t-p_2}|^\delta) + \log|\varepsilon_t|. \tag{3.3}$$

In the first step of the model selection procedure, $\log|a_t|$ is centered using the sample median, and the rest follows similarly to the conditional mean case. To achieve the parameter estimation using TLS, define $w_{q,a,t}$ to be the weight corresponding to the centered $\log|a_t|$ and $W_{q,a}$ to be the diagonal matrix with diagonal entries $(w_{q,a,p_1+p_2+1}, \ldots, w_{q,a,T})$. Also, let $\log|\boldsymbol{\varepsilon}| = (\log|\varepsilon_{p_1+p_2+1}|, \ldots, \log|\varepsilon_T|)'$. It is then assumed that $E[W_{q,a}\log|\boldsymbol{\varepsilon}|] = \mathbf{0}$ and $E[\log|\boldsymbol{\varepsilon}|' W_{q,a}\log|\boldsymbol{\varepsilon}|] < \infty$. Table 4 displays the number of time and lag parameters selected. Notice that time parameters play a role in predicting conditional volatility ($\max\{(t/T - 1/3)^3, 0\}$ for IBM and $\max\{(t/T - 2/3)^3, 0\}$ for HP), and that the selected model is considerably more parsimonious than the full model.

Data	δ
IBM	0.34
HP	0.29
SAP	0.31

Table 5: The selected δ for which S_q is minimized for $|\varepsilon_t|^\delta$.

3.3 Modeling the multiplicative error

It remains to determine an appropriate structure for the multiplicative errors in the conditional volatility model, whose distribution plays a crucial role in quantile estimation. The tail behavior of ε_t requires a particularly careful treatment as the VaR is typically calculated at the lower 1st or 5th percentile level. The most popular distributional assumption for ε_t, such as the one used in *RiskMetrics*™ (Longerstaey and More, 1995), is the standard normal distribution. However, due to empirical evidence of heavy-tailedness, Student's t-distribution or the generalized error distribution have been utilized as alternatives (see Tsay, 2010, Chapter 7).

The motivation for proposing the (three-parameter) generalized gamma distribution to fit $|\varepsilon_t|$ is as follows. It is known that the cube-root transformation of a gamma variable makes the underlying distribution roughly symmetric if the shape parameter is not too small (see Krishnamoorthy, Mathew, and Mukherjee (2008) for more discussion). Now, a preliminary study finds that the normal distribution fits adequately to $|\varepsilon_t|^{1/3}$ in the center but not in the tails. Table 5 suggests that $\delta \approx 1/3$ also approximately symmetrizes the distribution of $|\varepsilon_t|^\delta$ across different stocks, which is consistent with the results found for $|a_t|^\delta$. This leads to an initial guess that a gamma distribution may be able to approximate the distribution of $|\varepsilon_t|$. To improve the tail behavior, the generalized gamma distribution is used to model the multiplicative error. Its probability density function is given by

$$p(x;\boldsymbol{\theta}) = \frac{1}{\Gamma(\ell)} c a^{\ell c} x^{\ell c - 1} \exp\{-(ax)^c\}, \quad x > 0, \tag{3.4}$$

for $\boldsymbol{\theta} = (a, \ell, c)'$, where $a > 0$, $\ell > 0$, $c > 0$, and $\Gamma(z) = \int_0^\infty e^{-x} x^{z-1} dx$ is the Gamma

Data	a	ℓ	c
IBM	1.19	2.02	2.49
HP	1.42	2.55	2.06
SAP	1.24	2.15	2.35

Table 6: Generalized gamma parameter estimates for the truncated $|\varepsilon_t|^{1/3}$.

function (Lawless, 1980; Gomes, Combes, and Dussauchoy, 2008). To satisfy the assumption $E[\log|\varepsilon_t|] = \mathbf{0}$ and to estimate parameters in the conditional volatility model reliably, in the following, the generalized gamma distribution is fitted to the truncated $|\varepsilon_t|^{1/3}$ using the corresponding weights $w_{q,a,t} = 1$. Suppose further that ξ is a random variable with probability density function (3.4), then

$$E[\log \xi] = \frac{\psi(\ell)}{c} - \log a,$$

where $\psi(z) = \Gamma'(z)/\Gamma(z)$ is the digamma function (Lawless, 1980). Hence, the assumption $E[\log \xi] = 0$, corresponding to the mean of the truncated $|\varepsilon_t|$ being equal to zero, implies the relationship

$$a = \exp\left(\frac{\psi(\ell)}{c}\right) \tag{3.5}$$

for the generalized gamma parameters. Using $c = 3$ (corresponding to the cube-root transformation assuming $|\varepsilon_t|$ is gamma distributed) and the method-of-moment estimate for ℓ (see pg. 958 of Gomes, Combes, and Dussauchoy, 2008) as initial estimates, maximum likelihood estimates are obtained for the truncated $|\varepsilon_t|^{1/3}$, utilizing the constraint (3.5) for the parameter a (see Table 6).

Table 7 shows the comparisons between the percentiles of the original and truncated $|\varepsilon_t|^{1/3}$ as well as the percentiles of the fitted generalized gamma distribution. The generalized gamma distribution approximated the distribution of both the original and truncated $|\varepsilon_t|^{1/3}$ well, including the upper tails. It should therefore be useful for volatility predictions.

Original $\lvert\varepsilon_t\rvert^{1/3}$							
Data	**2%**	**10%**	**25%**	**50%**	**75%**	**90%**	**98%**
IBM	0.37	0.61	0.82	1.05	1.24	1.43	1.69
HP	0.37	0.59	0.81	1.03	1.27	1.46	1.74
SAP	0.33	0.60	0.83	1.03	1.26	1.45	1.72
Truncated $\lvert\varepsilon_t\rvert^{1/3}$							
Data	**2%**	**10%**	**25%**	**50%**	**75%**	**90%**	**98%**
IBM	0.44	0.65	0.83	1.06	1.24	1.43	1.69
HP	0.43	0.63	0.82	1.04	1.28	1.47	1.74
SAP	0.44	0.65	0.84	1.04	1.26	1.46	1.74
Estimated Generalized Gamma Distribution							
Data	**2%**	**10%**	**25%**	**50%**	**75%**	**90%**	**98%**
IBM	0.46	0.66	0.83	1.04	1.25	1.45	1.71
HP	0.45	0.64	0.82	1.04	1.27	1.49	1.78
SAP	0.45	0.65	0.83	1.04	1.26	1.47	1.73

Table 7: Percentile comparisons.

4 Predictive performance

In this section, the prediction performance of the proposed conditional volatility model is compared against three popular models, namely the GARCH(1,1) model of Bollerslev (1986), the APGARCH(1,1) of Ding, Granger and Engle (1993), and a simple nonstationary volatility model similar to Fryzlewicz, Sapatinas and Subba Rao (2006) that specifies a time trend only. This type of model provides a good benchmark as it was shown to have competitive predictive performance in the latter paper. Although it is possible to compare to GARCH(p, q) (or APGARCH(p, q)) with higher orders, estimation procedures for these cases are not necessarily numerically stable. Moreover, model selection criteria that would aid in automatically determining a specific order (p, q) to use for the forecast are not well established. For the evaluation, n-step-ahead predictions were computed with n varying from 1 to 5.

Letting $a_t = \sigma_t \phi_t$ where ϕ_t are i.i.d. random variables with mean 0 and variance 1, the GARCH(1,1) model imposes the structure $\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2$, on the conditional volatility, where $\alpha_0 > 0$, $\alpha_1 \geq 0$ and $\beta_1 \geq 0$. The APGARCH(1,1) model uses the form $\sigma_t^\lambda = \alpha_0 + \alpha_1(|\phi_{t-1}| - \gamma_1 \phi_{t-1})^\lambda a_{t-1}^2 + \beta_1 \sigma_{t-1}^\lambda$, where $\alpha_0 > 0$, $\alpha_1 \geq 0$, $\beta_1 \geq 0$, $\lambda > 0$, and $-1 < \gamma_1 < 1$. The parameters of both GARCH(1,1) and APGARCH(1,1) models are obtained by quasi-maximum likelihood estimation (QMLE) assuming standard normally distributed ϕ_t. Consistency and asymptotic normality of the QMLE parameters are documented in Berkes, Horváth and Kokoszka (2003) and Penzer, Wang and Yao (2009).

The simple nonstationary volatility model that specifies a time trend only is (3.2) without any lagged response variables. That is, $\log \sigma_t = h_0(t/T)$ where $h_0(t/T)$ is modeled with cubic splines. Predictors are included in the model with the selection procedure M_{W_q} and parameter estimation is done using TLS with $q = 99$. The mean function is modeled as outlined in Section 2.2.

Let v_{T+n} be the n-step-ahead prediction of the (appropriately scaled) conditional volatility. For the GARCH(1,1) and the APGARCH(1,1) model, these predictions are based on the conditional expectation $E[|a_{T+n}|^\delta | \mathcal{F}_T]$ with $\delta = 2$ and $\delta = \lambda$, respectively.

It follows then that

$$v_{T+n} = (E[|a_{T+n}|^{\delta}|\mathcal{F}_T])^{1/\delta}.$$

The sample counterpart, for which the parameters are estimated by QMLE. is denoted by $\hat{v}_{T+n}$.

The prediction procedure for the proposed model is as follows. First, assuming that $|\varepsilon_t|^{1/3}$ follows the generalized gamma distribution with parameter vector $\boldsymbol{\theta} = (a, \ell, c)'$ as in Section 3.3, one calculates recursively $\tilde{\sigma}_{T+i} = \exp\{E[\log \sigma_{T+i}|\mathcal{F}_T]\}, \quad i = 1, \ldots, n$, using the regression specified by (3.2) and (3.3) with plug-in estimates $\tilde{\sigma}_{T+1}, \ldots, \tilde{\sigma}_{T+i-1}$ in case $i \neq 1$. It follows, in particular, that

$$E\big[|a_{T+i}|^{1/3}|\mathcal{F}_T\big] = \tilde{\sigma}_{T+i}^{1/3} E\big[|\varepsilon_t|^{1/3}\big] = \frac{\tilde{\sigma}_{T+i}^{1/3}\Gamma(\ell + \frac{1}{c})}{a\Gamma(\ell)}$$

for $i = 1, \ldots, n-1$. The n-step-ahead predictor is then given by

$$v_{T+n} = \tilde{\sigma}_{T+n}\big(E\big[|\varepsilon_t|^{1/3}\big]\big)^3 = \tilde{\sigma}_{T+n}\left[\frac{\Gamma(\ell + \frac{1}{c})}{a\Gamma(\ell)}\right]^3$$

and its sample counterpart $\hat{v}_{T+n}$, for which the parameters $\boldsymbol{\theta} = (a, \ell, c)'$ are estimated by maximum likelihood estimation based on the truncated $|\varepsilon_t|^{1/3}$ (see Section 3.3) and the coefficients of the model by TLS (see Section 2.1). Since the distribution of $|\varepsilon_t|^{1/3}$ is approximately symmetric, it follows that $(E[|\varepsilon_t|^{1/3}])^3 \approx \text{median}(|\varepsilon_t|)$. To measure the accuracy of the n-step-ahead prediction $\hat{v}_{T+n}$, the n-step-ahead prediction of the conditional mean is estimated first. This is given by $\mu_{T+n} = E[r_{T+n}|\mathcal{F}_T]$ and its sample counterpart denoted by $\hat{\mu}_{T+n}$ in which the coefficients of the model are estimated by TLS, noting that $E[a_{T+n}|\mathcal{F}_T] = 0$. Now, define the absolute prediction error

$$\text{APE}_{n,T} = \big||\hat{a}_{T+n}| - \hat{v}_{T+n}\big|$$

where $\hat{a}_{T+n} = r_{T+n} - \hat{\mu}_{T+n}$.

The prediction performance is evaluated using the sliding window approach, where the training set of $T = 1099$ returns is varied 500 times. In particular, the first training set is from Friday, April 5, 2002 to Tuesday, August 15, 2006, the second training set is from Monday, April 8, 2002 to Wednesday, August 16, 2006, and the last (500th) training set is from Monday, March 29, 2004 to Friday, August 8, 2008.

It is examined how the absolute prediction errors compare locally by plotting a smoothed version, obtained from 5-point moving averages, of the annualized $\text{APE}_{n,T}$ ($\text{APE}_{n,T}$ multiplied by 250). The results for $n = 2$ are displayed in Figure 3. It can be seen that the annualized $\text{APE}_{n,T}$ of the proposed model are lower than those of the competitor models in the majority of cases.

The overall performance is evaluated by comparing the proportion of these 500 $\text{APE}_{n,T}$ obtained by the proposed model higher than those of the competitor models. Two measures of proportion (Proportion 1 and Proportion 2) are reported. Proportion 1 is defined as the proportion of instances for which the proposed method had a higher absolute prediction error than the corresponding competitors. Proportion 2 is defined as the probability that a randomly chosen absolute prediction error obtained from the proposed method is larger than a randomly selected absolute prediction error obtained from all prediction errors (proposed and competitor methods combined). For Proportion 2, the Munzel (1999) nonparametric paired two-sample test which also utilizes the proportion, or equivalently for the two-sample case, the *relative treatment effect* denoted by RTE, is carried out. For the Munzel test, the alternative hypothesis is one-sided stating that the RTE of the proposed model is lower than that of the competitors, or equivalently, that Proportion 2 is less than 0.5. By using this approach, the comparisons can be made in a robust manner.

The results in Table 8 indicate that the proposed model performs favorably compared to the GARCH(1,1) and APGARCH(1,1) models overall for one up to five steps ahead predictions. This is especially the case for SAP where the $\text{APE}_{n,T}$ of the proposed model is higher than the GARCH-type models only roughly 1/3 of time when each pair is compared separately. When the proposed model is compared to the time trend only model, interestingly, there was no significant difference for one-step ahead predictions, but significant differences were observed for multiple-step ahead predictions. In addition to the proportions, Table 9 reports the median of these 500 annualized $\text{APE}_{n,T}$ for each n, denoted by $\text{MAPE}_{n,T}$. The results are in line with the findings in Table 8.

$\text{APE}_{n,T}$ Proportion Comparisons							
Data	**Model Compared**	**Statistic**	$n=1$	$n=2$	$n=3$	$n=4$	$n=5$
IBM	Time Trend Only	Proportion 1	0.51	0.23	0.23	0.24	0.25
		Proportion 2	0.50	0.41	0.42	0.42	0.42
		p-value	*0.62*	$\mathbf{< 10^{-4}}$	$\mathbf{< 10^{-4}}$	$\mathbf{< 10^{-4}}$	$\mathbf{< 10^{-4}}$
	GARCH(1,1)	Proportion 1	0.43	0.44	0.43	0.44	0.44
		Proportion 2	0.46	0.46	0.46	0.46	0.47
		p-value	$\mathbf{< 10^{-4}}$	$\mathbf{< 10^{-4}}$	$\mathbf{< 10^{-4}}$	$\mathbf{< 10^{-4}}$	$\mathbf{< 10^{-4}}$
	APGARCH(1,1)	Proportion 1	0.44	0.43	0.45	0.44	0.46
		Proportion 2	0.47	0.47	0.47	0.47	0.47
		p-value	$\mathbf{< 10^{-4}}$	$\mathbf{< 10^{-4}}$	$\mathbf{< 10^{-4}}$	$\mathbf{< 10^{-4}}$	$\mathbf{< 10^{-4}}$
HP	Time Trend Only	Proportion 1	0.46	0.28	0.28	0.29	0.28
		Proportion 2	0.50	0.44	0.45	0.45	0.45
		p-value	*0.05*	$\mathbf{< 10^{-4}}$	$\mathbf{< 10^{-4}}$	$\mathbf{< 10^{-4}}$	$\mathbf{< 10^{-4}}$
	GARCH(1,1)	Proportion 1	0.39	0.39	0.37	0.38	0.37
		Proportion 2	0.43	0.43	0.43	0.44	0.43
		p-value	$\mathbf{< 10^{-4}}$	$\mathbf{< 10^{-4}}$	$\mathbf{< 10^{-4}}$	$\mathbf{< 10^{-4}}$	$\mathbf{< 10^{-4}}$
	APGARCH(1,1)	Proportion 1	0.42	0.40	0.42	0.40	0.41
		Proportion 2	0.45	0.45	0.45	0.45	0.45
		p-value	$\mathbf{< 10^{-4}}$	$\mathbf{< 10^{-4}}$	$\mathbf{< 10^{-4}}$	$\mathbf{< 10^{-4}}$	$\mathbf{< 10^{-4}}$
SAP	Time Trend Only	Proportion 1	0.48	0.26	0.26	0.27	0.26
		Proportion 2	0.50	0.43	0.42	0.43	0.43
		p-value	*0.27*	$\mathbf{< 10^{-4}}$	$\mathbf{< 10^{-4}}$	$\mathbf{< 10^{-4}}$	$\mathbf{< 10^{-4}}$
	GARCH(1,1)	Proportion 1	0.32	0.32	0.32	0.32	0.31
		Proportion 2	0.43	0.42	0.42	0.42	0.41
		p-value	$\mathbf{< 10^{-4}}$	$\mathbf{< 10^{-4}}$	$\mathbf{< 10^{-4}}$	$\mathbf{< 10^{-4}}$	$\mathbf{< 10^{-4}}$
	APGARCH(1,1)	Proportion 1	0.35	0.34	0.33	0.32	0.32
		Proportion 2	0.43	0.43	0.43	0.42	0.42
		p-value	$\mathbf{< 10^{-4}}$	$\mathbf{< 10^{-4}}$	$\mathbf{< 10^{-4}}$	$\mathbf{< 10^{-4}}$	$\mathbf{< 10^{-4}}$

Table 8: Proportions of $\text{APE}_{n,T}$ by the proposed model higher than those of the time trend only model, GARCH(1,1), and APGARCH(1,1), and the p-values obtained from the Munzel test (*Italics*: non-significant, **Bold**: significant). For all models, the data are detrended a priori using the conditional mean modeling approach of Section 2.

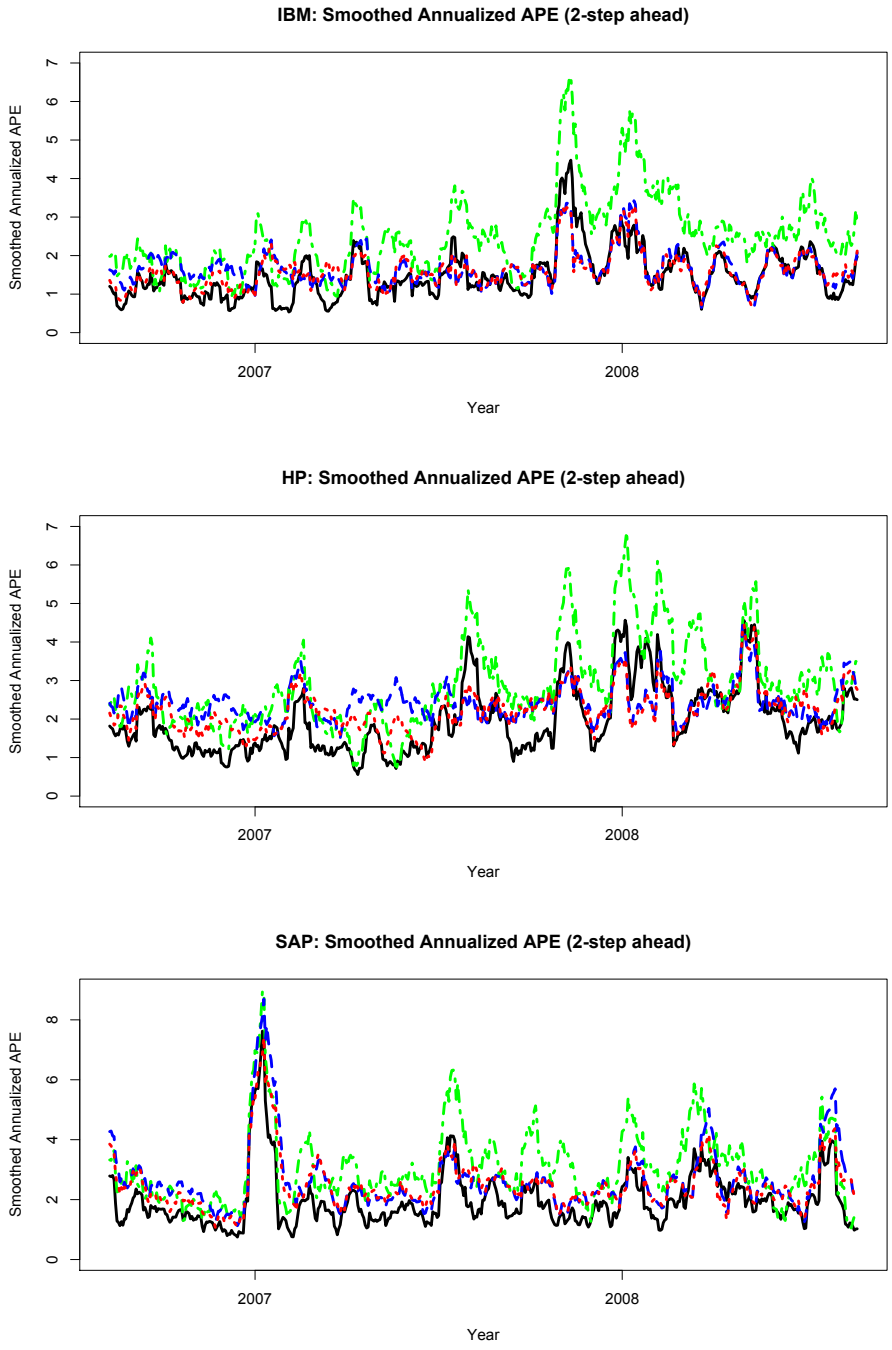

Figure 3: Smoothed annualized $\mathrm{APE}_{n,T}$ for the 2-step-ahead predictions ($n = 2$). The black solid line indicates the proposed model. The green broken line indicates the time trend only model. The blue dashed line indicates GARCH(1,1) and the red dotted line indicates APGARCH(1,1).

Annualized $\text{MAPE}_{n,T}$						
Data	**Model**	$n=1$	$n=2$	$n=3$	$n=4$	$n=5$
IBM	Proposed Model	1.04	**1.05**	**1.02**	**1.05**	**1.05**
	Time Trend Only	**1.03**	2.02	1.98	1.99	2.00
	GARCH(1,1)	1.51	1.53	1.51	1.47	1.49
	APGARCH(1,1)	1.42	1.36	1.38	1.40	1.42
HP	Proposed Model	**1.42**	**1.47**	**1.46**	**1.52**	**1.42**
	Time Trend Only	1.44	2.08	2.06	2.13	2.09
	GARCH(1,1)	2.32	2.38	2.39	2.43	2.44
	APGARCH(1,1)	2.11	2.12	2.20	2.22	2.19
SAP	Proposed Model	**1.32**	**1.29**	**1.34**	**1.38**	**1.29**
	Time Trend Only	1.43	2.26	2.27	2.18	2.15
	GARCH(1,1)	2.36	2.44	2.38	2.41	2.39
	APGARCH(1,1)	2.29	2.28	2.35	2.32	2.34

Table 9: Comparisons of annualized $\text{MAPE}_{n,T}$ ($250\text{MAPE}_{n,T}$). For the GARCH(1,1) and APGARCH(1,1) models, the data are detrended *a priori* using the conditional mean modeling approach of Section 2. The best ones are indicated in **bold**.

5 Rates of convergence for spline-based estimators

In this section, some theoretical results on large-sample properties of the proposed model are provided in a simplified setting. Note that mean and volatility processes may be nonstationary and are of the form $Y_t = g_0(t/T) + g_1(Y_{t-1}) + \cdots + g_p(Y_{t-p}) + Z_t$, where g_0 is a smooth function of (rescaled) time, $g_1, \ldots, g_p$ smooth functions of lagged response variables, and Z_t an innovation term with mean zero and variance σ_Z^2. In the following, the case $p = 1$ is considered. The arbitrary p case can be treated with similar arguments. Let

$$Y_t = m(t) + Z_t, \qquad m(t) = g_0(t/T) + g_1(X_t),$$

with $X_t = Y_{t-1}$. It is assumed that the X_t take values in a compact set, taken to be $[0, 1]$ without loss of generality. For identifiability, it is required that $\int_0^1 g_1(u)du = 0$. Moreover, the following smoothness conditions on g_0 and g_1 are needed.

Assumption 5.1. *Let g_0 and g_1 be η times differentiable such that the ηth derivatives of g_0 and g_1 satisfy*

$$|g_0^{(\eta)}(u_1) - g_0^{(\eta)}(u_2)| \leq K|u_1 - u_2|^\zeta, \qquad |g_1^{(\eta)}(u_1) - g_1^{(\eta)}(u_2)| \leq K|u_1 - u_2|^\zeta$$

for $\zeta \in [0, 1)$ and $K > 0$. Denote $\nu = \eta + \zeta$.

The function of rescaled time g_0 is estimated by the spline estimator $\beta_0' B_0(t/T)$, where $B_0(u)$ is a vector of B-splines of degree $d \geq \nu$ with k equispaced knots in $[0, 1]$. Since $\int_0^1 g_1(u)du = 0$ by assumption, the spline estimator for g_1 needs to integrate to zero as well. This can be accomplished as in Burman (1991). Let therefore $B_1(x)$ be the vector of B-splines of degree d with k knots. The dimension of $B_1(x)$ is $k + 2d$. Let Ψ_1 be a $(k + 2d - 1) \times (k + 2d)$ matrix whose rows are of unit length, orthogonal to each other, and orthogonal to $\int_0^1 B_1(x)dx$. If $\Phi_1(x) = \Psi_1 B_1(x)$, then by construction $\int_0^1 \Phi_1(x)dx = 0$. The estimate of g_1 is consequently of the form $\beta_1' \Phi_1(x)$. Let $B_t = (B_0'(t/T), \Phi_1'(X_t))'$ and $\beta = (\beta_0', \beta_1')'$. Combining the preceding lines, the estimate of $m_t = g_0(t/T) + g_1(X_t)$ is therefore of the form $\beta' B_t$. An estimate of β can

be obtained by minimizing the least squares criterion

$$\frac{1}{T}\sum_{t=1}^{T}(Y_t - \beta' B_t)^2$$

with respect to β. With the solution

$$\hat{\beta} = \left(\frac{1}{T}\sum_{t=1}^{T} B_t B_t'\right)^{-1} \frac{1}{T}\sum_{t=1}^{T} B_t Y_t,$$

the estimator for m_t becomes $\hat{m}_t = \hat{\beta}' B_t$. The rate of convergence of this spline-based estimator can be established if the following conditions are met.

Assumption 5.2. *Assume that the following requirements hold.*

(a) There are constants $0 < \underline{f} < \overline{f}$ *such that* $\underline{f} \le f_t(x) \le \overline{f}$ *for all* t *and* x*, where* f_t *is the probability density function of* X_t*.*

(b) The series (X_t) *is* α*-mixing, that is,*

$$\big|P(A \cap B) - P(A)P(B)\big| \le \alpha(s),$$

for all $A \in \sigma(\ldots, X_{t-1}, X_t)$ *and* $B \in \sigma(X_{t+s}, X_{t+s+1}, \ldots)$*, with* $\sum_{s=0}^{\infty} \alpha(s) < \infty$*.*

The main result on the rate of convergence is formulated as a theorem.

Theorem 5.1. *If Assumptions 5.1 and 5.2 hold and if* $k^3/T \to 0$*, then the spline-based estimator* $\hat{m}_t$ *satisfies*

$$\frac{1}{T}\sum_{t=1}^{T}(\hat{m}_t - m_t)^2 = \mathcal{O}_P\big(T^{-2\nu/(2\nu+1)}\big) \qquad (T \to \infty).$$

In the proof of the theorem, given in the online supplement Noguchi, Aue and Burman (2014), it is assumed that (i) the knots in estimating g_0 and g_1 are equispaced, and (ii) the number of knots in estimating g_0 and g_1 are the same, namely k. As pointed out in Stone (1985), these are simplifying assumptions that can be suitably modified. The proofs for the general case remains the same but requires increasingly complicated notations.

The condition $k^3/T \to 0$ can be improved further by requiring different mixing conditions. More details on this are provided in the proof of Lemma 2 in the online

supplement. The rate given in Theorem 5.1 is the optimal rate of convergence for estimating a univariate function nonparametrically, see Stone (1977). More specifically,

$$\frac{1}{T}\sum_{t=1}^{T}(\hat{m}_t - m_t)^2 = \mathcal{O}_P\Big(\frac{k}{T}\Big) + \mathcal{O}_P\big(k^{-2\nu}\big),$$

so that the minimum is attained when k is of the order $T^{1/(2\nu+1)}$, which in turn implies the statement of the theorem. Set $D = T^{-1}\sum_{t=1}^{T} B_t B_t'$ and let $\bar{\beta} = D^{-1}T^{-1}\sum_{t=1}^{T} B_t m_t$ and $\bar{m}_t = \bar{\beta}' B_t$. The method of proof will depend on showing that

$$\frac{1}{T}\sum_{t=1}^{T}(\hat{m}_t - \bar{m}_t)^2 = \mathcal{O}_P\Big(\frac{k}{T}\Big) \qquad \text{and} \qquad \frac{1}{T}\sum_{t=1}^{T}(\bar{m}_t - m_t)^2 = \mathcal{O}_P\big(k^{-2\nu}\big).$$

6 Conclusions

In this paper, a new financial time series model has been proposed. Cubic splines have been employed to model the conditional mean and conditional volatility jointly, including time and lagged response variables as predictors. The model can be written as a linear regression, thereby making predictions possible in a simple manner. The (three-parameter) generalized gamma distribution has been determined as a suitable distribution for the multiplicative error of the conditional volatility. For both the absolute residual $|a_t|$ and multiplicative error $|\varepsilon_t|$, it has been observed that the cube-root transformation makes the underlying distribution symmetric. A suitable information criterion has been suggested to perform model selection. The overall predictive performance of the proposed model using three individual daily stock returns compares favorably to time trend only, GARCH(1,1) and APGARCH(1,1) predictions.

As an extension of the model considered in this paper, one may apply the generalized linear model as the generalized gamma distribution is found to be appropriate for the multiplicative error. Further large-sample properties, such as those of the maximum likelihood estimators of the parameters using the generalized linear model approach, may be addressed in future work.

References

[1] Akaike, H. (1974), "A New Look at the Statistical Model Identification," *IEEE Transactions on Automatic Control*, 19, 716–722.

[2] Aue, A., Berkes, I., and Horváth, L. (2006), "Strong Approximation for the Sums of Squares of Augmented GARCH Sequences," *Bernoulli*, 12, 583–608.

[3] Berkes, I., Horváth, L., and Kokoszka, P. (2003), "GARCH Processes: Structure and Estimation," *Bernoulli*, 9, 201–227.

[4] Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307–327.

[5] Burman, P. (1991), "Regression Function Estimation from Dependent Observations," *Journal of Multivariate Analysis*, 36, 263–279.

[6] Burman, P., and Nolan, D. (1995), "A General Akaike-type Criterion for Model Selection in Robust Regression," *Biometrika*, 82, 877–886.

[7] Dahlhaus, R., and Subba Rao, S. (2006), "Statistical Inference for Time-varying ARCH Processes," *The Annals of Statistics*, 34, 1075–1114.

[8] Ding, Z., Granger, C., and Engle, R.F. (1993), "A Long Memory Property of Stock Market Returns and a New Model," *Journal of Empirical Finance*, 1, 83–106.

[9] Engle, R.F. (1982), "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987–1007.

[10] Engle, R.F. (2002), "New Frontiers for ARCH Models," *Journal of Applied Econometrics*, 17, 425–446.

[11] Engle, R.F., and Rangel, J.G. (2008), "The Spline-GARCH Model for Low-frequency Volatility and its Global Macroeconomic Causes," *The Review of Financial Studies*, 21, 1187–1222.

[12] Franke, J., Härdle, W., and Hafner, C.M. (2008), *Statistics of Financial Markets: An Introduction*, Berlin: Springer-Verlag.

[13] Fryzlewicz, P., Sapatinas, T., and Subba Rao, S. (2006), "A Haar–Fisz Technique for Locally Stationary Volatility Estimation," *Biometrika*, 93, 687–704.

[14] Gao, J., and Tong, H. (2004), "Semiparametric Non-linear Time Series Model Selection," *Journal of the Royal Statistical Society, Series B*, 66, 321–336.

[15] Glosten, L., Jagannathan, R., and Runkle, D. (1993), Relationship Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks," *The Journal of Finance*, 48, 1779–1801.

[16] Gomes, O., Combes, C., and Dussauchoy, A. (2008), "Parameter Estimation of the Generalized Gamma Distribution," *Mathematics and Computers in Simulation*, 79, 955–963.

[17] Kim, W., and Linton, O. (2004), "The LIVE Method for Generalized Additive Volatility Models," *Econometric Theory*, 20, 1094–1139.

[18] Krishnamoorthy, K., Mathew, T., and Mukherjee, S. (2008), "Normal-based Methods for a Gamma Distribution," *Technometrics*, 50, 69–78.

[19] Lawless, J.F. (1980), "Inference in the Generalized Gamma and Log Gamma Distributions," *Technometrics*, 22, 409–419.

[20] Linton, O., and Nielsen, J.P. (1995), "A Kernel Method of Estimating Structured Nonparametric Regression Based on Marginal Integration," *Biometrika*, 82, 93–100.

[21] Longerstaey, J., and More, L. (1995), *Introduction to RiskMetrics*, Fourth Edition. New York: Morgan Guaranty and Trust Company.

[22] Mikosch, T., and Stărică, C. (2004), "Stock Market Risk-Return Inference. An Unconditional Non-parametric Approach," Preprint, University of Copenhagen. Available at: `http://www.math.ku.dk/ mikosch/Preprint/Catalin/jf020904.pdf`.

[23] Munzel, U. (1999), "Nonparametric Methods for Paired Samples," *Statistica Neerlandica*, 53, 277–286.

[24] Nelson, D.B. (1991), "Conditional Heteroscedasticity in Asset Returns: a New Approach," *Econometrica*, 59, 347–370.

[25] Noguchi, K., Aue, A. and Burman, P. (2014). Online supplementary material to "Exploratory analysis and modeling of stock returns". Available at `http://...`.

[26] Penzer, J., Wang, M., and Yao, Q. (2009), "Approximating Volatilities by Asymmetric Power GARCH Functions," *Australian and New Zealand Journal of Statistics*, 51, 201–225.

[27] Rech, G., Teräsvirta, T., and Tschernig, R. (2001), "A Simple Variable Selection Technique for Nonlinear Models," *Communications in Statistics—Theory and Methods*, 30, 1227–1241.

[28] Stone, C.J. (1977), "Consistent Nonparametric Regression," *The Annals of Statistics*, 5, 595–645.

[29] Stone, C.J. (1985), "Additive Regression and Other Nonparametric Models," *The Annals of Statistics*, 13, 689–705.

[30] Tsay, R.S. (2010), *Analysis of Financial Time Series*, Third Edition. New York: Wiley.

[31] Tschernig, R., and and Yang, L. (1999), "Nonparametric Lag Selection for Time Series," *Journal of Time Series Analysis*, 21, 457–487.

[32] Vogt, M. (2012), "Nonparametric Regression for Locally Stationary Time Series," *The Annals of Statistics*, 40, 2601–2633.

[33] Yang, L., Härdle, W., and Nielsen, J.P. (1999), "Nonparametric Autoregression with Multiplicative Volatility and Additive Mean," *Journal of Time Series Analysis*, 20, 579–604.