Kocaeli Üniversitesi **Sosyal Bilimler** Enstitüsü **Dergisi** (14) 2007 / 2 : 78-109

The Use of ARCH and GARCH Models for Estimating and Forecasting Volatility

Bahadtin Rüzgar*
İsmet Kale**

Abstract: This paper presents the performance of 11 ARCH-type models each with four different distributions combined with ARMA specifications in conditional mean in estimating and forecasting the volatility of IMKB 100 stock indices, using daily data over a 9 years period. The results suggest that fractionally integrated asymmetric models outperform the non-FI versions and, using skewed-t and student-t distributions provide better fit to the data for almost every model in estimating volatility. In forecasting volatility a clear improvement is not observed by altering a specific model component or distribution.

Keywords: GARCH; EGARCH; GJR; APARCH; IGARCH; FIGARCH; FIAPARCH; FIEGARCH; HYGARCH; ARMA; GED; Skewed-t; Ox; G@RCH

1. Introduction

Until the 80's most of the analytical research was on finding the relation between factors and outcomes. For this purpose it was a mostly used assumption for simplicity that errors were random constants. The models that best represent relations were those that produced minimum errors. As a result the errors were minimized in models, and remained out of the subject of quantitative prediction. However in some cases it is exactly the quantity of those errors, the prediction of which is important. The world of finance is one of the first who supported the research as they realized that this error can be interpreted as what we may call the risk.

Generalized AutoRegressiv Conditional Heteroskedasticity (GARCH) is a model of errors. It is mostly used in other models to represent volatility. The models that

---

* Yrd. Doç. Dr. Bahadtin Rüzgar, Marmara Üniversitesi Bankacılık ve Sigortacılık Yüksekokulu Aktüerya Bölümü'n'de öğretim üyesidir.
** İsmet Kale, Marmara Üniversitesi Bankacılık ve Sigortacılık Yüksekokulu Bankacılık Bölümü'nde Yüksek Lisans öğrencisidir.

make use of GARCH vary from predicting the spread of toxic gases in the atmosphere to simulating neural activity. But finance is still the leading area and dominates the research on GARCH.

ARCH class models were first introduced by Nobel price awarded Engle (1982) with the ARCH model. Since then, numerous extensions have been put forward, all of them modelling the conditional variance as a function of past (squared) returns and associated characteristics.

In recent years, the tremendous growth of trading activity and trading losses of financial institutions has led financial regulators and supervisory committee of banks to favor quantitative techniques which appraise the possible loss that these institutions can incur. Value-at-Risk (VaR) has become one of the most sought-after techniques. The computation of the VaR for a collection of returns requires the computation of the empirical quantile at level α of the distribution of the returns of the portfolio. Because quantiles are direct functions of the variance in parametric models, ARCH class models immediately translate into conditional VaR models. These conditional VaR models are important for characterizing short term risk for intradaily or daily trading positions.

In this paper we investigate the estimating and forecasting capabilities of GARCH models when applied to daily IMKB 100 index data. We furthermore aim to understand whether IMKB data exhibits the common caracteristics of financial time series observed in developed countries. We thereby wish to contribute to the risk management research in Turkey, the outcomes of which will be of crucial value after the implementation of Basel II regulations in 2007.

The rest of the paper is organized in the following way. In Section 2, we describe ARMA and GARCH processes as the building blocks of analysed variance models. These models are applied to daily stock index data in Section 3 where we assess their performances and conclude.

2. Formation of variance models

GARCH models are designed to capture certain characteristics that are commonly associated with financial time series: fat tails, volatility clustering and leverage effects.

Probability distributions for asset returns often exhibit fatter tails than the standard normal, or Gaussian, distribution. Time series that exhibit a fat tail distribution are often referred to as leptokurtic. In addition, financial time series usually exhibit a characteristic known as volatility clustering, in which large changes tend to follow large changes, and small changes tend to follow small changes. In either case, the changes from one period to the next are typically of unpredictable sign. Large disturbances, positive or negative, become part of the information set used to construct the variance forecast of the next period's disturbance. In this manner, large shocks of either sign are allowed to persist, and can influence the volatility forecasts for several periods. Volatility clustering, or persistence, suggests a time-series model in which successive disturbances, although uncorrelated, are nonetheless serially dependent.

Finally, certain classes of asymmetric GARCH models are also capable of capturing the so-called leverage effect, in which asset returns are often observed to be negatively correlated with changes in volatility.

A standard approach of time series analysis is to take a time series that exhibits complicated behavior and try to convert it to a simpler form. Optimally, such simplification would yield time series that were so simple that they could reasonably be modeled as independent and identically distributed (IID). In practice, and especially in financial applications, this is rarely possible. Stationarity is a condition similar to IID, but not as strong. Two different forms of stationarity are defined:

i) A process is said to be **strictly stationary** if the unconditional joint distribution of any segment $(y_t, y_{t+1}, ..., y_{t+r})$ is identical to the unconditional joint distribution of any other segment $(y_{t+s}, y_{t+s+1}, ..., y_{t+s+r})$ of the same length.

ii) A process is said to be **covariance stationary** if the unconditional joint distribution of any segment $(y_t, y_{t+1}, ..., y_{t+r})$ has means, standard deviations and correlations that are identical to the corresponding means, standard deviations and correlations of the unconditional joint distribution of any other segment $(y_{t+s}, y_{t+s+1}, ..., y_{t+s+r})$ of equal length. Correlations include autocorrelations and cross correlations.

Strict stationarity is appealing because it affords a form of homogeneity across terms without requiring that they be independent. Covariance stationarity is the condition that is more frequently assumed in GARCH models. It does require that all first and second moments exist whereas strict stationarity does not. In this one respect, covariance stationarity is a stronger condition (Holton, 1996).

In this paper we are going to construct linear models combining mean and variance equations holding either covariance or strict stationary. We will use ARMA for mean and GARCH for variance spesification.

2.1 ARMA (*R*, *S*) And ARFIMA (*R*, *D*, *S*) Processes In The Conditional Mean

Box and Jenkins introduced a flexible family of time series models capable of expressing a variety of short-range serial relationships in terms of linear regression, where the predictors are previous observations and previous residual errors.

One component of the Box-Jenkins framework is the autoregressive (AR) equation, where predicting variables are previous observations. An AR(r) model, where the predictors are the previous r terms, is defined as

$$y_t = \sum_{i=1}^{r} \phi_i y_{t-i} + \varepsilon_t \tag{1}$$

In equation 1, the current value y_t is partly based on the value at time t - i ($i \leq r$), and partly based on a random variable ε, typically Gaussian noise. The influence of prior values is usually assumed to decay over time, such that $\phi_1 > \phi_2 > \ldots > \phi_r$.

A second component in the Box-Jenkins framework is the moving average (MA) process. In an MA process, the observation y_t is dependent not on the previous values of y_t, but rather on the values of the noise random variable ε. A moving average model of order s, MA(s), is defined by

$$y_t = \sum_{j=1}^{s} \theta_j \varepsilon_{t-j} + \varepsilon_j \tag{2}$$

where y_t depends on the previous s errors ε_{t-j} ($j \leq s$) and the current error ε_t.
Ooms and Doornik (1999) present the basic ARMA(r,s) model as

$$y_t = \phi_1 y_{t-1} + \ldots + \phi_r y_{t-r} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_s \varepsilon_{t-s} \tag{3}$$

and the general ARMA (r,s) is in the form

$$y_t = \phi_0 + \sum_{i=1}^{r} \phi_i y_{t-i} + \varepsilon_t + \sum_{j=1}^{s} \phi_j \varepsilon_{j-i} \tag{4}$$

where r is the order of the AR(r) part, ϕ_i its parameters, s the order of the MA(s) part, θ_j its parameters and ε_t normally and identically distributed noise or innovation process.

The family of ARMA models as defined by equation 4 is flexible and able to concisely describe the serial dependencies of seemingly complex time series in terms of the number of parameters (i.e., the order or history) of the AR and MA components, and the values of these parameters.

In fields such as physics and economics, phenomena that fluctuate over time often display long-range serial correlations. In order to correctly identify and parsimoniously describe processes that give rise to persistent serial correlations, traditional ARMA time series models can be extended to allow for fractional integration to capture long-range correlations. The resulting ARFIMA models, popular in econometrics and hydrology, allow for simultaneous maximum likelihood estimation of the parameters of both short-range and long-range processes.

Following the description of Laurent and Peters (2002), by using lag polynominals and introducing a mean μ, the equation 4 becomes

$$\phi(L)(y_t - \mu) = \theta(L)\varepsilon_t \tag{5}$$

where L is the lag operator, μ *is* the unconditional mean of y_t,

$\phi(L) = 1 - \sum_{i=1}^{r} \phi_i L^i$ and $\theta(L) = 1 + \sum_{i=1}^{s} \theta_i L^i$ are the autoregressiv and the moving average operators in the lag operator. They are polinominals of order r and s respectively. With a fractional integration parameter d, the ARFIMA(r, d, s) model is written as

$$\phi(L)(1-L)^d (y_t - \mu) = \theta(L)\varepsilon_t \tag{6}$$

The fractional differencing operator $(1\text{-}L)^d$ is a notation for the following infinite polynominal:

$$(1-L)^d = \sum_{i=0}^{\infty} \frac{\Gamma(i-d)}{\Gamma(i+1)\Gamma(-d)} L^i \equiv \sum_{i=0}^{\infty} \pi_i(d) L^i \tag{7}$$

where $\pi_i(z) \equiv \Gamma(i-z)/\Gamma(i+1)\Gamma(-z)$ and $\Gamma(.)$ is the Standard gamma function. To ensure statitionary and invertibility of the process y_t, d lies between -0.5 and 0.5. Given data series y_t one can use conditional or exact likelyhood method to specify the order and parameters. The Ljung-Box statistics of residuals can check the fit.

Bhardwaj and Swanson (2004) found that ARFIMA models perform better for greater forecast horizons and that they under certain conditions provide significantly better out-of-sample predictions than AR, MA, ARMA, GARCH, simple regime switching, and related models.

Throughout the paper ARMA spesification will only be used to model the mean of returns. ARMA (0,0) implies a constant mean, ARMA(1,0) is simply AR(1). It is also possible to make the conditional mean a function of the conditional variance. In that case the conditional variance derived from the GARCH model will be a variable in the mean equation. This then will be the so called ARCH-in-mean model, which we denote in this paper with (-m) in naming our models.

2.2 GARCH Processes to Model The Conditional Variance

If the value of the stock market index at time t is marked P_t, the return of the index at time t is given by $y_t = \ln(P_t / P_{t-1})$ where ln denotes natural logarithm.

For the log return series y_t, we assume its mean is ARMA modelled, then let $\varepsilon_t = y_t - \mu_t$ be the mean corrected log return. Stock market index returns can be modelled with the help of the following equation:

$$y_t = \mu + \varepsilon_t, \tag{8}$$

where μ is the mean value of the return, which is expected to be zero; t is a random component of the model, not autocorrelated in time, with a zero mean value. Sequence ε_t may be considered a stochastic process, expressed as:

$$\varepsilon_t = z_t \ \sigma_t \tag{9}$$

where z_t is a sequence of independently and identically distributed random variables, with a distribution $E(z_t) = 0$ and $Var(z_t) = 1$. By definition ε_t is serially uncorrelated with a mean equal to zero, but its conditional variance equals σ_t^2 and therefore may change over time, contrary to what is assumed in the standard regression model. The conditional variance is the measure of our uncertainty about a variable given a model and information set.

Following Markowitz' definition of volatility as standard deviation of the expected return, σ_t is the volatility of log returns at time t, the changes of which will be modelled by means of the following ARCH-type models.

2.2.1 The ARCH Model

Volatility clustering, or persistence, suggests a time-series model in which successive disturbances, although uncorrelated, are nonetheless serially dependent. Rob Engle had the great insight to introduce and study the class of autoregressive conditionally heteroscedastic (ARCH) time series models for modeling the time-varying volatility clustering phenomenon (Engle, 1982). He used a weighted average of squared past residuals over a long period with higher weights on the recent past and small but non-zero weights on the distant past.

The ARCH (q) model can be expressed as

$$\varepsilon_t = z_t \sigma_t \tag{10}$$

$$z_t \sim i.i.d\ D(0,1)$$

$$\sigma_t^2 = \sigma^2 (\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots, t, x_t, b) = \sigma^2 (\sigma_{t-1} z_{t-1}, \sigma_{t-2} z_{t-2}, \ldots, t, x_t, b)$$

where ε_t denotes the prediction error at time t, x_t is a vector of lagged exogenous variables, b is a vector of parameters, D(.) is distribution.

The conditional variance of ε_t given the information at time t-1 is σ_t^2. For the parameterization of this variance many possibilities are suggested in the literature. In its original form ARCH can be written as

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \ldots + \alpha_q \varepsilon_{t-q}^2 \tag{11}$$

using $\sum$ operator equation 11 becomes

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 \tag{12}$$

or by replacing $\varepsilon_t = z_t \sigma_t$, to more clearly notice the autoregression, we get

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \sigma_{t-i}^2 z_{t-i}^2 \tag{13}$$

The ARCH model can describe volatility clustering through the following mechanism: if ε_{t-1} was large in absolute value, σ_t^2 and thus ε_t is expected to be large in absolute value as well. Even if the conditional variance of an ARCH model is time varying, the unconditional variance of ε_t is constant provided that $\alpha_0 > 0$ and $\sum_{i=1}^{q} \alpha_i < 1$.

Conditional variance σ_t^2 has to be positive for all t. Sufficient conditions are when $\alpha_0 > 0$ and $\alpha_i \geq 0$. Evidence has shown that a high ARCH order has to be selected to catch the dynamics of the conditional variance. This involves the estimation of a large number of parameters. The generalized ARCH (GARCH) model of Bollerslev (1986) is based on an infinite ARCH spesification and it allows reducing the number of estimated parameters by imposing nonlinear restrictions on them.

2.2.2 The GARCH Model

The GARCH model additionally assumes that forecasts of variance changing in time also depend on the lagged conditional variances of capital assets. An unexpected increase or fall in the returns of an asset at time t will generate an increase in the variability expected in the period to come.

Introduced by Engle (1982) and Bollerslev (1986) the mostly used GARCH (p,q) models make σ_t^2 a linear function of lagged conditional variances and squared past residual

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \ldots + \alpha_q \varepsilon_{t-q}^2 + \beta_1 \sigma_{t-1}^2 + \ldots + \beta_p \sigma_{t-p}^2 \tag{14}$$

using $\sum$ operator

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_i \sigma_{t-i}^2 \tag{15}$$

where p is the degree of GARCH; q is the degree of the ARCH process, $\alpha_0 >0$, $\alpha_i \geq 0$, $\beta_j \geq 0$. The covariance stationary condition is $\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1$. Since the equation expresses the dependence of the variability of returns in the current period on data (i.e. the values of the variables ε_{t-i}^2 and σ_{t-j}^2) from previous periods, we denote this variability as conditional.

One can observe that an important feature of the GARCH (p,q) model is that it can be regarded as an ARMA (r,s), where r is the larger of p and q. This result allows econometricians to apply the analysis of ARMA process to the GARCH model.

Using the lag operator, the GARCH (p,q) model can be rewritten as: ($\omega=\alpha_0$)

$$\sigma_t^2 = \omega + \alpha(L)\,\varepsilon_t^2 + \beta(L)\,\sigma_t^2 \qquad (16)$$

where L denotes the lag operator and $\alpha(L)$ and $\beta(L)$ denote the AR and MA polynominals respectively, with $\alpha(L) = \alpha_1(L) + \alpha_2(L)^2 + \ldots + \alpha_q(L)^q$ and $\beta(L) = \beta_1(L) + \beta_2(L)^2 + \ldots + \beta_p(L)^p$.

If all the roots of the polynominal $|1- \beta(L)| =0$ lie outside of the unit circle, we get:

$$\sigma_t^2 = \omega\,|\,1- \beta(L)\,|^{-1} + \alpha(L)\,|\,1- \beta(L)\,|^{-1}\,\varepsilon_t^2, \qquad (17)$$

which may be regarded as an ARCH (∞) process, since the conditional variance linearly depends on all previous squared residuals. The unconditional variance is given by:

$$\sigma^2 \equiv E(\varepsilon_t^2) = \frac{\omega}{1 - \sum_{i=1}^{q} \alpha_i - \sum_{j=1}^{p} \beta_j} \qquad (18)$$

The basic and most widespread model is GARCH (1, 1), which can be reduced to:

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \qquad (19)$$

As the variance is expected to be positive, we expect that the regression coefficients ω, α, β are always positive (α and β can also take the value 0), while the stationarity of the variance is preserved, if the the sum of α and β is smaller than 1. Conditional variability of the returns defined in equation 19 is determined by three effects:

1. The constant part, which is given by the coefficient ω;

2. Part of variance expressed by the relationship $\alpha\varepsilon_{t-1}^2$ and designated as ARCH component;

3. Part given by the predicted variability from the previous period and expressed by the relationship $\beta\sigma_{t-1}^2$.

The sum of regression coefficients ($\alpha+\beta$) expresses the influence of the variability of variables from the previous period on the current value of the variability. This value is usually close to 1, which is a sign of increased effects of shocks on the variability of returns on financial assets.

While the basic GARCH model allows a certain amount of leptokurtic behaviour this is often insufficient to explain real world data. We therefore use 3 distributions other than normal in our analysis, namely Student-t, Skewed-t (Lambert and Laurent, 2000-2001) and Generalized Error Distributions which help to allow for the fat tails in the distribution.

The choice of the quadratic form for the conditional variance has the important consequence that the impact of the past values of the innovation on the current volatility is only a function of their magnitude not of its sign. The principal disadvantage of the GARCH model is therefore its unsuitability or modelling the frequently observed asymmetry that occurs when a different volatility is recorded systematically in the case of good and bad news.

Falls and increases in the returns can be interpreted as good and bad news. If a fall in returns is accompanied by an increase in volatility greater than the volatility induced by an increase in returns, we may speak of a 'leverage effect'. Following classes of asymmetric GARCH models are capable of capturing this effect.

2.2.3 The GJR and TARCH model

The GJR model is an asymmetric model. It is proposed by Glosten, Jagannathan and Runkle (1993). The generalized version may be written as

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_1 \varepsilon_{t-1}^2 + \sum_{i=1}^{q} \gamma_i S_{t-i} \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \tag{20}$$

where S_t is a dummy variable with $S_{t-i}=1$, if $\varepsilon_{t-i} < 0$ and $S_{t-i}=0$, if $\varepsilon_{t-i} \geq 0$.

In this model, it is assumed that the impact of ε_t^2 on the conditional variance σ_t^2 is different when ε_t is positive or negative. The TARCH model of Zakoian (1994) is very similar to the GJR, where he preferred to model the standard deviation instead of the conditional variance. Its basic variant is GJR (1,1), which is expressed by:

$$\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \gamma \sigma_{t-1}^2 + \beta \varepsilon_{t-1}^2 S_{t-1} \tag{21}$$

The model can be interpreted that unexpected (unforeseen) changes in the returns of the index y_t expressed in terms of ε_t, have different effects on the conditional variance of stock market index returns. An unforeseen increase is presented as good news and contributes to the variance in the model through multiplicator α. An unforeseen fall, which is a bad news, generates an increase in volatility through multiplicators α and β. The asymmetric nature of the returns is then given by the nonzero value of the coefficient β, while a positive value of β indicates a 'leverage effect'. The covariance stationary condition is $\sum_{t=1}^{q} \alpha_i (1+\gamma_i^2) + \sum_{j=1}^{p} \beta_j < 1$.

2.2.4 The EARCH Model

The exponential GARCH (EGARCH) model is introduced by Nelson (1991). In this model, the conditional variance may be expressed as follows:

$$\ln \sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i s(z_{t-1}) + \sum_{j=1}^{p} \beta_j \ln(\sigma_{t-j}^2) \tag{22}$$

where $z_t = \varepsilon_t / \sigma_t$ is the normalized residuals series. The function $s(.)$ can be written as

$$s(z_t) = \delta_1 z_t + \delta_2 \{ |z_t| - E(|z_t|) \} \tag{23}$$

Therefore $\delta_1 z_t$ adds the effect of the sign of ε_t whereas δ_2 {$|z_t|$ - E ($|z_t|$) adds its magnitude effect. E ($|z_t|$) depends on the choice of the distribution of return series. For the normal distribution E ($|z_t|$) = $\sqrt{\frac{2}{\pi}}$.

Its basic variant is EGARCH (1, 1) with normal distribution is expressed by:

$$\ln \sigma_t^2 = \omega + \alpha (\delta_1 \frac{\varepsilon_{t-1}}{\sigma_{t-1}} + \delta_2 \left[\left| \frac{\varepsilon_{t-1}}{\sigma_{t-1}} \right| - \sqrt{\frac{2}{\pi}} \right]) + \beta \ln \sigma_{t-1}^2 \tag{24}$$

The asymmetric nature of the returns is then given by the nonzero value of the coefficient δ_1, while a positive value of δ_1 indicates a 'leverage effect '.

The use of ln transformation ensures that σ_t^2 is always positive and consequently there are no restrictions on the sign of the parameters. Moreover external unexpected shocks will have a stronger influence on the predicted volatility than TARCH or GJR.

2.2.5 The APARCH Model

In general, the inclusion of a power term acts so as to emphasise the periods of relative tranquillity and volatility by magnifying the outliers in that series. Squared terms are therefore so often used in models. If a data series is normally distributed than we are able to completely characterise its distribution by its first two moments (McKenzie and Mitchell, 2001). If we accept that the data may have a non-normal error distribution, other power transformations may be more appropriate.

Recognising the possibility that a squared power term may not necessarily be optimal, Ding, Granger and Engle (1993) introduced a new class of ARCH model called the Power ARCH (PARCH) model. Rather than imposing a structure on the data, the Power ARCH class of models estimates the optimal power term.

Ding, Granger and Engle (1993) also specified a generalised asymmetric version of the Power ARCH model (APARCH). The APARCH (p,q) model can be expressed as:

$$\sigma_t^\delta = \alpha_0 + \sum_{i=1}^{q} \alpha_i (|\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i})^\delta + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^\delta \tag{25}$$

where $\alpha_0 > 0$, $\delta \geq 0$, $\beta_j \geq 0$, $\alpha_i \geq 0$ and $-1 < \gamma_i < 1$.

This model couples the flexibility of a varying exponent δ with the asymmetry coefficient γ_i to take the "leverage effect" into account. Moreover, the APARCH includes ARCH, GARCH and GJR as special cases:

ARCH when $\delta = 2$, $\gamma_i = 0$ (i = 1,. . . ,p) and $\beta_j = 0$ (j = 1,. . . ,p),

GARCH when $\delta = 2$ and $\gamma_i = 0$ (i = 1,. . . ,p) and

GJR when $\delta = 2$

It also includes four other ARCH extentions which are not tested in this paper

TARCH when $\delta = 1$,

NARCH when $\gamma_i = 0$ (i = 1,. . . ,p) and $\beta_j = 0$ (j = 1,. . . ,p),

The Log-ARCH, when $\delta \to 0$ and

Taylor / Schwert GARCH when $\delta = 1$, and $\gamma_i = 0$ (i = 1, . . . , p).

A stationary solution for APARCH model exists. See Ding, Granger and Engle (1993) for details.

2.2.6 The IGARCH Model

In explaining the GARCH (p, q) model it was mentioned that GARCH may be regarded as an ARCH (∞) process, since the conditional variance linearly depends on all previous squared residuals. Moreover it was stated that a GARCH process is covariance stationary if and only if $\sum_{i=1}^{q}\alpha_i + \sum_{j=1}^{p}\beta_j < 1$. But strict stationarity does not require such a stringent restriction that the unconditional variance does not depend on t, in fact we often find in estimation that $\sum_{i=1}^{q}\alpha_i + \sum_{j=1}^{p}\beta_j$ is close to 1.

Lets denote h as the timelag between the present shock and future conditional variance. Then a shock to the conditional variance σ_t^2 has a decaying impact on σ_{t+h}^2. When h increases this impact becomes neglectable indicating a short memory. However if $\sum_{i=1}^{q}\alpha_i + \sum_{j=1}^{p}\beta_j \geq 1$, the effect on σ_{t+h}^2 does not die out even for a very high h. This property is called persistence in the literature. In many high frequency time series applications, the conditional variance estimated using GARCH (p,q) process exhibits a strong persistence.

It was also mentioned that the GARCH (p, q) process can be seen as an ARMA process. It is known that such an ARMA process has a unit root when $\sum_{i=1}^{q}\alpha_i + \sum_{j=1}^{p}\beta_j = 1$. When the sum of all AR coefficients and MA coefficients is equal to one, the ARMA process is integrated (ARIMA). Due to their similarity to ARMA models GARCH models are symetric and have short memory.

A GARCH model that satisfies $\sum_{i=1}^{q}\alpha_i + \sum_{j=1}^{p}\beta_j = 1$ (or equally rewritten as $\alpha(L)+\beta(L)=1$) is known as an integrated GARCH (IGARCH) process (Engle and Bollerslev, 1986), meaning that current information remains of importance when forecasting the volatility for all horizons. IGARCH (p, q) models are a kind of ARIMA models for volatilities. Recall that GARCH model is:

$$define\ \ \upsilon_t \equiv \varepsilon^2{}_t - \sigma^2{}_t$$

$$\sigma_t^2 = \omega + \alpha(L)\varepsilon^2 + \beta(L)\sigma_t^2 \quad (26)$$

$$\varepsilon^2{}_t = \omega + (\alpha(L)+\beta(L))\varepsilon^2 + \beta(L)\upsilon_t + \upsilon_t \quad (27)$$

This is an ARMA {max (p, q), p} model for the squared innovations. If $\alpha(L)+\beta(L)=1$ then we have an Integrated GARCH model (IGARCH).

The IGARCH can be expressed as:

$$\phi(L)(1-L)\varepsilon_t^2 = \omega + [1-\beta(L)]v_t \text{ or} \quad (28)$$

$$\phi(L)(1-L)\varepsilon_t^2 = \omega + [1-\beta(L)](\varepsilon_t^2 - \sigma_t^2), \quad (29)$$

where $\theta(L) = [1-\alpha(L)-\beta(L)](1-L)^{-1}$ is a polynominal of order {max (p,q)-1 }

We can also express the conditional variance as a function of the squared residuals, and then an IGARCH (p, q) becomes:

$$\sigma_t^2 = \frac{\omega}{[1-\beta(L)]} + \left\{1-\phi(L)(1-L)[1-\beta(L)]^{-1}\right\}\varepsilon_t^2 \quad (30)$$

Although a process y_t that follows an IGARCH process is not covariance stationary, and its unconditional variance is infinite, IGARCH process is still important since the unconditional density of y_t is the same for all t, and thus the process can come from a strictly stationary process. However we may suspect that IGARCH is more a product of omitted structural breaks than the result of true IGARCH behavior. An integrated process will be hereafter denoted as I(1), a non-integrared process I(0).

The assumption of short memory such as in GARCH models is usually not fulfilled. Ding Granger and Engle (1993) during their research for the APARCH model have found that the absolute returns and their power transformations have a highly significant long-term memory property as the returns are highly correlated. For example, significant positive autocorrelations were found at over 2,700 lags in 17,054 daily observations of the S&P 500. That makes 2700 lags / 252 trading days = 10.7 years. On the other hand the implications of IGARCH models are too strong which leads to the consideration of fractionally integrated models.

2.2.7 The FIGARCH Model

As shown in Ding, Granger, and Engle (1993) among others, the effects of a shock can take a considerable time to decay. Therefore, the distinction between I(0) and I(1) processes seems to be far too restrictive. In an I(0) process the propagation of shocks occurs at an exponential rate of decay so that it only captures the short-memory, while for an I(1) process the persistence of shocks is infinite.

In the conditional mean, the ARFIMA specification has been proposed to fill the gap between short and complete persistence, so that the short-run behavior of the time-series is captured by the ARMA parameters, while the fractional differencing parameter allows for modelling the long-run dependence.

The first long memory GARCH model was the fractionally integrated GARCH (FIGARCH) introduced by Ballie, Bollerslev and Mikkelsen (1996). The FIGARCH (p, d, q) model is a generalization of the IGARCH model by replacing the operator (1-L) of the IGARCH equation by $(1-L)^d$, where d is the memory parameter.

$$\phi(L)\left[(1-\alpha)+\alpha(1-L)^d\right]\varepsilon_t^2=\omega+\left[1-\beta(L)\right]\nu_t \qquad (31)$$

where $\theta(L)=\left[1-\alpha(L)-\beta(L)\right](1-L)^{-1}$ is a polynominal of order {max (p,q)-1 }(same as IGARCH). We can also express the conditional variance as a function of the squared residuals, then a FIGARCH (p,d,q) becomes:

$$\sigma_t^2=\frac{\omega}{\left[1-\beta(L)\right]}+\left\{1-\phi(L)(1-L)^d\left[1-\beta(L)\right]^{-1}\right\}\varepsilon_t^2 \qquad (32)$$

FIGARCH models exhibit long memory. They include GARCH models (for d=1) and IGARCH models (for d=1). In contrast to ARFIMA models, where the memory parameter d is $-0.5 < d < +0.5$, FIGARCH d is $0 < d < 1$.

FIGARCH-processes are non-stationary like IGARCH-processes. This shows that the concept of unit roots can hardly be generalized from linear to nonlinear processes. Furthermore, the interpretation of the memory parameter d is difficult in the FIGARCH set up.

2.2.8 The HYGARCH Model

Davidson (2001) extended the class of FIGARCH models to HYGARCH(p,α,d,q) models which stands for hyperbolic GARCH. HYGARCH-models replace the operator $(1-L)^d$ in FIGARCH equation by $[(1-\alpha)+\alpha(1-L)^d]$. The parametrization of HYGARCH-models is given by

$$\sigma_t^2 = \frac{\omega}{\left[1-\beta(L)\right]} + \left\{1-\left[1-\beta(L)\right]^{-1}\phi(L)\left\{1+\alpha\left[(1-L)^d\right]\right\}\right\}\varepsilon_t^2 \qquad (33)$$

The parameters α and d are assumed to be non-negative. HYGARCH-models nest GARCH models (for $\alpha = 0$), FIGARCH-processes (for $\alpha = 1$) and IGARCH-models (for $\alpha = d = 1$).

2.2.9 FI process of Chung v.s. Ballie, Bollerslev and Mikkelsen (BBM)

Chung (1999) underscores some little drawbacks in the BBM model: there is a structural problem in the BBM specification since the direct implementation of the ARFIMA framework originally designed for the conditional mean equation is not perfect for the use in conditional variance equation, leading to difficult interpretations of the estimated parameters.

Indeed the fractional differencing operator applies to the constant term in the mean equation (ARFIMA) while it does not in the variance equation (FIGARCH). Chung (1999) proposes a slightly different process:

$$\sigma_t^2 = \sigma^2 + \left\{1-\left[1-\beta(L)\right]^{-1}\phi(L)(1-L)^d\right\}(\varepsilon_t^2-\sigma^2) \qquad (34)$$

or

$$\sigma_t^2 = \sigma^2 + \lambda(L)(\varepsilon_t^2-\sigma^2) \qquad (35)$$

λ (L) is an infinite summation which, in practice, has to be truncated. BBM propose to truncate λ (L) at 1000 lags and initialize the unobserved ε_t^2 at their unconditional moment. Contrary to BBM, Chung (1999) proposes to truncate λ (L) at the size of the information set (t-1) and to initialize the unobserved (ε_t^2- σ^2) at 0. In our analysis we hold the proposal of BBM and truncate at 1000 lags.

2.2.10 The FIEGARCH and FIAPARCH Model

The idea of fractional integration has been extended to other GARCH types of models, including the Fractionally Integrated EGARCH (FIEGARCH) of Bollerslev and Mikkelsen (1996) and the Fractionally Integrated APARCH (FIAPARCH) of Tse (1998).

Similarly to the GARCH (p, q) process, the EGARCH (p, q) can be extended to account for long memory by factorizing the autoregressive polynomial $\left[1-\beta(L)\right] = \phi(L)(1-L)^d$ where all the roots of $\phi(z) = 0$ lie outside the unit circle. The FIEGARCH (p, d, q) is specified as follows:

$$\ln(\sigma_t^2) = \omega + \phi(L)^{-1}(1-L)^{-d}\left[1+\alpha(L)\right]s(z_{t-1}) \qquad (36)$$

And the FIAPARCH (*p, d, q*) model can be written as:

$$\sigma_t^{\delta} = \omega + \left\{1 - \left[1 - \beta(L)\right]^{-1} \phi(L)(1-L)^d \right\}(|\varepsilon_t| - \gamma\varepsilon_t)^{\delta} \tag{37}$$

3. Empirical Applications

In order to study the estimation and forecasting performances of different GARCH processes, 11 models are applied to Istanbul Stock Exchange 100 Index log returns. We used daily data from 3.1.1996 to 15.12.2004. From a total of 2200 trading days 2000 are used for estimation and 200 are left to test the forecasts.

The models applied are GARCH, EGARCH, GJR, APARCH, IGARCH, FIGARCH of BBM, FIGARCH of Chung, FIEGARCH, FIAPARCH of BBM, FIAPARCH of Chung and HYGARCH. The (p,q) = (1,1) variant of all models are systematically tested with four different distributions, namely, Gaussian Normal, Student-t, Generalized Error Distribution (GED) and Skewed-t distribution. That makes 44 basic models. Moreover for matters of observation 50 more models of higher order combined with ARMA are unsystematically experimented to examine the effects on estimation and forecasting performances with an emphasis on Maximum Likelihood.

We have written an Ox code named IMKB_Estimate&Forecast.ox on the basis of the examples and objects provided in the Ox 3.40 and its G@RCH 3.00 package and used it for our analysis. G@RCH 3.0 of Laurent and Peters (2002) is a package dedicated to GARCH models and many of its extensions. It is written in the Ox programming language (see Doornik, 1999). G@RCH 3.0 can be downloaded free of charge for academic purposes at http://www.egss.ulg.ac.be/garch/.

The program proved to be very flexible and fast. Currently only sharewares provided for academic research are capable of analyzing such recent variety of models. Moreover most standard software do not allow for such a flexible combination of processes like we applied in our analysis. With open-source code are able to add or modify specifications, processes or graphics in the future.

Estimation results are evaluated on the basis of ML, Akaike, Schwarz, Shibata and Hannan-Quinn values, whereas forecasting results are ranked according to Mincer Zarnowitz regression R^2, Root Mean Square Error, Mean Square Error and Mean Absolute Error value criteria.

3.1 Estimation Results

It is apparent that t distributions shall be preferred if one aims to obtain a better representation of the existing data. Among the first 15 best basic estimating models according to all five criteria all was either student-t or skewed-t. GED and Normal distributions follow. Therefore in comparing the estimation powers of models, we restrict our comments to student and skewed-t distributions. Test statistics of some models in estimating performances and forecasting performances according to different criteria are given Table 3.1.5 and Table 3.2.5.

Table 3.1.1: First 20 And Last Ten Models With Constant Mean İn Estimating Performances According To Different Criteria

	Log-L	Akaike $-2\frac{LogL}{n}+2\frac{k}{n}$	Schwartz $-2\frac{LogL}{n}+2\frac{k\log[\log(n)]}{n}$	Shibata $-2\frac{LogL}{n}+2\frac{\log(k)}{n}$	Hannan-Quinn $-2\frac{LogL}{n}+\log\left(\frac{n+2k}{n}\right)$	
5	FIAparchCh11Skt	FIAparchBBM11St-t	FIgarchCh11St-t	FIAparchBBM11St-t	FIgarchCh11St-t	5
	FIAparchBBM11Skt	FIAparchCh11Skt	FIgarchBBM11St-t	FIAparchCh11Skt	FIgarchBBM11St-t	
	FIAparchBBM11St-t	FIAparchBBM11Skt	HYGarch11St-t	FIAparchBBM11Skt	FIAparchBBM11St-t	
	FIAparchCh11St-t	FIAparchCh11St-t	FIgarchCh11Skt	FIAparchCh11St-t	FIAparchCh11St-t	
	HYGarch11Skt	FIgarchCh11St-t	FIGarchBBM11Skt	FIgarchCh11St-t	HYGarch11St-t	
10	HYGarch11St-t	FIgarchBBM11St-t	FIAparchBBM11St-t	FIgarchBBM11St-t	FIgarchCh11Skt	10
	FIgarchCh11Skt	HYGarch11St-t	Igarch11St-t	HYGarch11St-t	FIGarchBBM11Skt	
	FIGarchBBM11Skt	FIgarchCh11Skt	FIAparchCh11St-t	FIgarchCh11Skt	FIAparchCh11Skt	
	FIgarchCh11St-t	FIGarchBBM11Skt	Garch11St-t	FIGarchBBM11Skt	FIAparchBBM11Skt	
	FIgarchBBM11St-t	HYGarch11Skt	Gjr11St-t	HYGarch11Skt	HYGarch11Skt	
15	Aparch11Skt	Gjr11St-t	HYGarch11Skt	Gjr11St-t	Gjr11St-t	15
	Gjr11Skt	Gjr11Skt	FIAparchCh11Skt	Gjr11Skt	Gjr11Skt	
	Aparch11St-t	Aparch11St-t	FIAparchBBM11Skt	Aparch11St-t	Garch11St-t	
	Gjr11St-t	Aparch11Skt	Igarch11Skt	Aparch11Skt	Aparch11St-t	
	FIAparchCh11GED	Garch11St-t	Garch11Skt	Garch11St-t	Igarch11St-t	
20	FIAparchBBM11GED	Garch11Skt	Gjr11Skt	Garch11Skt	Garch11Skt	20
	Garch11Skt	FIAparchCh11GED	FIGarchCh11GED	FIAparchCh11GED	Aparch11Skt	
	HYGarch11GED	FIAparchBBM11GED	Aparch11St-t	FIAparchBBM11GED	FIGarchCh11GED	
	Garch11St-t	FIGarchCh11GED	FIGarchBBM11GED	FIGarchCh11GED	FIGarchBBM11GED	
	FIGarchCh11GED	FIGarchBBM11GED	Aparch11Skt	FIGarchBBM11GED	IGarch11Skt	
	……………	……………	……………	……………	……………	
10	FIgarchBBM11N	HYGarch11N	Garch11N	HYGarch11N	HYGarch11N	10
	EGarch11Skt	Gjr11N	Igarch11N	Gjr11N	Garch11N	
	Aparch11N	Garch11N	Gjr11N	Garch11N	Gjr11N	
	Gjr11N	Aparch11N	Aparch11N	Aparch11N	Aparch11N	
	Garch11N	EGarch11Skt	EGarch11Skt	EGarch11Skt	Igarch11N	
5	Igarch11N	Igarch11N	EGarch11GED	Igarch11N	EGarch11Skt	5
	EGarch11GED	EGarch11GED	FIEgarch11N	EGarch11GED	EGarch11GED	
	FIEgarch11St-t	FIEgarch11St-t	FIEgarch11St-t	FIEgarch11St-t	FIEgarch11N	
	FIEgarch11N	FIEgarch11N	EGarch11N	FIEgarch11N	FIEgarch11St-t	
	EGarch11N	EGarch11N	FIAparchBBM11GED	EGarch11N	EGarch11N	

Log-L = log likelihood value, n = number of observations, k = number of estimated parameters

For optimizing maximum likelihood, skewed-t performs better than student-t for all models. On the other hand if we evaluate according to the other four criteria, by which more complicated models are penalized for the inclusion of additional parameters, skewed-t looses its apparent advantage, because it requires an additional skewness parameter. Especially Hannan-Quinn test seems to judge according to the distribution rather than the model specification and prefer student-t.

We found that the choice of models is at least as important as the choice of distributions, because best performing models combined with both distributions found place in the front ranks, mostly successively. For ranking models with constant mean in estimating performances according to different criteria, test statistics, illustrated on Table 3.1.5, are used.

Table 3.1.2: Minimum Sum Of Rankings Of Different GARCH And FIAPARCH Specifications

	Log- L	Akaike	Shibata	Hannan-Quinn	Total
AR1Garch22Skt	2	1	1	2	6
Garch22Skt	3	2	2	1	8
ARMA22Garch-m33Skt	1	3	3	7	14
AR1Garch-m11Skt	4	4	4	6	18
AR1Garch11Skt	5	5	5	5	20
Garch11St-t	7	6	6	3	22
Garch11Skt	6	7	7	4	24
Garch11GED	8	8	8	8	32
AR1Garch11N	9	9	9	9	36
Garch11N	10	10	10	10	40

……………………………………………………………………………………

	Log-L	Akaike	Schwarz	Shibata	Hannan-Quinn	Total
AR1FIAparchCh11Skt	3	2	3	2	3	13
ARMA11FIAparchCh21Skt	1	1	7	1	6	16
AR1FIAparchCh21Skt	2	4	5	3	5	19
AR1FIAparchCh21St-t	4	3	4	4	4	19
FIAparchCh11Skt	5	5	2	5	2	19
FIAparchCh11St-t	6	6	1	6	1	20
FIAparchCh11GED	8	7	6	7	7	35
AR1FIAparchCh21GED	7	8	8	8	8	39
AR1FIAparchCh21N	9	9	10	9	10	47
FIAparchCh11N	10	10	9	10	9	48

Log Likelihood results are consistent with aggregate results. Normal distribution estimates worse for all models. Higher orders alone improve results more than ARMA specifications alone. Together they improve more but the marginal benefit decreases.

Among the basic models fractionally integrated ones, especially FIAPARCH of BBM and that of Chung combined with student-t and skewed-t distributions perform outstanding based on all criteria. It is also to note that based on the estimation power, the methods of BBM and Chung report only slight differences. We can conclude that among models with the same distribution and same mean specification a FI (1) model is a better estimator than its FI (0) counterpart. This is a clear indicator that IMKB 100 indices shows strong persistence and the effect of shocks influence future returns for long periods. That FI (1) performs better then I (1) show that the persistence is not completely permanent.

Table 3.1.3: Comparison Of APARCH And FIAPARCH Models

	Model Structure			
Model	Mean Equation	Variance Equation	Distribution	Log-L
ARMA11Aparch33Skt	ARMA(1,1)	Aparch (3,3)	Skewed-t	4239.96
AR1Aparch22Skt	ARMA(1,0)	Aparch (2,2)	Skewed-t	4238.40
Aparch22Skt	ARMA(0,0)	Aparch (2,2)	Skewed-t	4237.11
Aparch33Skt	ARMA(0,0)	Aparch (3,3)	Skewed-t	4236.68
AR1Aparch11Skt	ARMA(1,0)	Aparch (1,1)	Skewed-t	4230.44
Aparch11Skt	ARMA(0,0)	Aparch (1,1)	Skewed-t	4228.85
Aparch11St-t	ARMA(0,0)	Aparch (1,1)	Student-t	4227.98
ARMA11Aparch-m11GED	ARMA(1,1)	Aparch (1,1)	GED	4221.97
Aparch11GED	ARMA(0,0)	Aparch (1,1)	GED	4218.36
Aparch11N	ARMA(0,0)	Aparch (1,1)	Normal	4170.24
………				
AR1FIAparchBBM21Skt	ARMA(1,0)	FIAparchBBM (2,d,1)	Skewed-t	4239.05
AR1FIAparchBBM11St-t	ARMA(1,0)	FIAparchBBM (1,d,1)	Student-t	4238.10
FIAparchBBM11Skt	ARMA(0,0)	FIAparchBBM (1,d,1)	Skewed-t	4237.78
FIAparchBMM11St-t	ARMA(0,0)	FIAparchBBM (1,d,1)	Student-t	4236.83
AR1FIAparchBBM11GED	ARMA(1,0)	FIAparchBBM (1,d,1)	GED	4227.41
FIAparchBMM11GED	ARMA(0,0)	FIAparchBBM (1,d,1)	GED	4226.64
FIAparchBBM11N	ARMA(0,0)	FIAparchBBM (1,d,1)	Normal	4181.37

A ranking among the specifications would be the existence of FI followed by the existence of asymetry. The coefficients indicating asymmetry are mostly significant and showing that there exist a leverage effect in IMKB 100. The simple GARCH (1,1) with t distributions follow the more complex models with t distributions but is clearly better than any model, even the FI(1) and asymmetric models, combined with GED or Normal distribution. No EGARCH model could reach strong convergence during numerical optimization and results are misleading.

With distributions and mean specifications equal, the FI (1) version outperforms the FI (0) version based on Log Likelihood.

Given the same distribution and same GARCH orders, using an autoregression AR(1) or autoregression+moving average ARMA (1,1) or an ARCH-in-mean effect in the mean equation improves the performance of all models if the ranking is based on maximum likelihood. Again the other four criteria take the additional parameters in account, however most of the time the improvement in maximum likelihood is large enough to compensate for the estimation burden of additional parameters. The marginal improvement decreases as the mean equation gets more complex.

The stand-alone effect of increasing the orders is significantly more than standalone effect of manipulating the mean equation. Orders of (2, 2) or (3, 3) perform well, but the chances of non-convergence and getting misleading results also increases.

Table 3.1.4: Best 20 Estimating Models Based On Log Likelihood

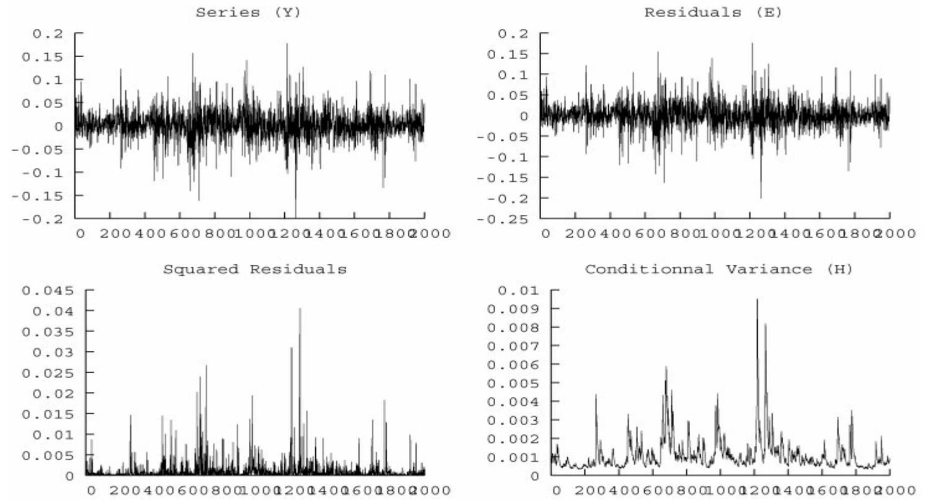
	Model Structure				
Model	Mean Equation	Variance Equation	Distribution	Log-L	R^2
ARMA11FIAparchCh21Skt	ARMA(1,1)	FIAparchBBM(2,d,1)	Skewed-t	4241.82	0.011570
AR1FIAparchCh21Skt	ARMA (1,0)	FIAparchCh (2,d,1)	Skewed-t	4239.99	0.012283
ARMA11Aparch33Skt	ARMA (1,1)	Aparch (3,3)	Skewed-t	4239.96	0.008243
ARMA11GJR33Skt	ARMA (1,1)	GJR (3,3)	Skewed-t	4239.43	0.020429
AR1FIAparchCh11Skt	ARMA (1,0)	FIAparchCh (1,d,1)	Skewed-t	4239.07	0.012136
AR1FIAparchBBM21Skt	ARMA (1,0)	FIAparchBBM(2,d,1)	Skewed-t	4239.05	0.011850
AR1FIAparchCh21St-t	ARMA (1,0)	FIAparchCh (2,d,1)	Student-t	4238.99	0.011743
AR1Gjr22Skt	ARMA (1,0)	GJR (2,2)	Skewed-t	4238.46	0.008727
ARMA22Garch-m33Skt	ARMA (2,2)	ARCH-mGARCH (3,3)	Skewed-t	4238.41	0.014413
AR1Aparch22Skt	ARMA (1,0)	Aparch (2,2)	Skewed-t	4238.40	0.008754
AR1FIAparchBBM11St-t	ARMA (1,0)	FIAparchBBM (1,d,1)	Student-t	4238.10	0.011746
FIAparchCh11Skt	ARMA (0,0)	FIAparchCh (1,d,1)	Skewed-t	4237.81	0.012526
FIAparchBBM11Skt	ARMA (0,0)	FIAparchBBM(1,d,1)	Skewed-t	4237.78	0.012499
Aparch22Skt	ARMA (0,0)	Aparch (2,2)	Skewed-t	4237.11	0.008964
FIAparchBBM11St-t	ARMA (0,0)	FIAparchBBM(1,d,1)	Student-t	4236.83	0.012130
Gjr33Skt	ARMA (0,0)	GJR (3,3)	Skewed-t	4236.68	0.021404
Aparch33Skt	ARMA (0,0)	Aparch (3,3)	Skewed-t	4236.68	0.021544
FIAparchCh11St-t	ARMA (0,0)	FIAparchCh (1,d,1)	Student-t	4236.47	0.008248
AR1HYGarch22Skt	ARMA (1,0)	HYGarch (2,d,2)	Skewed-t	4236.00	0.012553
AR1HYGarch11Skt	ARMA (1,0)	HYGarch (1,d,1)	Skewed-t	4235.85	0.011973

FIAPARCH models of different specifications and distributions dominate. Including mean specifications clearly improve Log-L results. Skewed-t distribution seems to be the solution for fat tails.

The combination of more complex mean equation with higher orders perform overall better with the cost of additional parameters. As a result we can conclude that, based on the maximum likelihood, the more complex a model is the better it fits to the data.

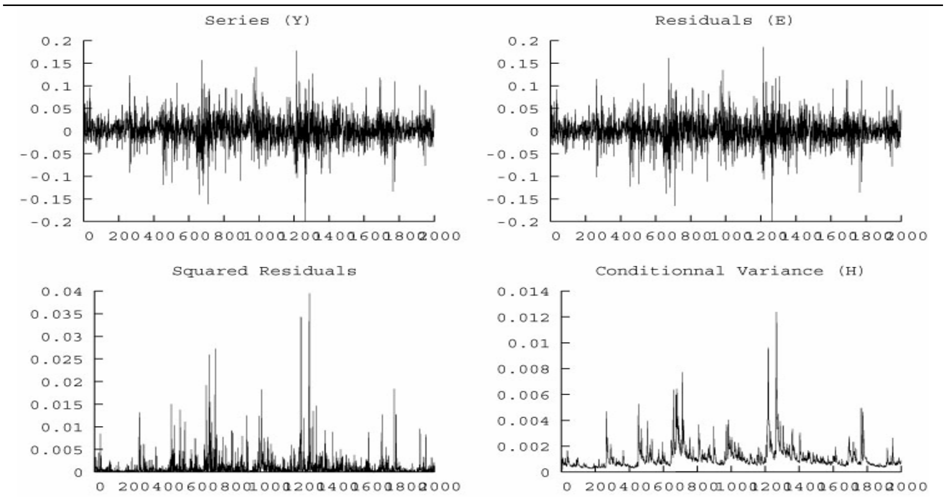
Combined rankings allow for the conclusion that maximum likelihood is a consistent evaluation criterion. While Akaike, Shibata, Schwarz and Hannan-Quinn may result in complete different rankings, the aggregate results are consistent with that of the maximum likelihood.

If skewed-t or student-t distributions are used, it is always possible to increase the likelihood by adding more tailored processes, implying a better fit to the data on the basis of numbers. However using the graphs we can show that even the simplest GARCH (1, 1) is a satisfactory model in estimation. The differences are not subtle and using a GARCH (1, 1) model would not lead to a different decision than a decision based on a more complex ARMA (1, 1) FIAPARCH (2, 1) model. GARCH models in general succeed in reproducing volatility clustering, persistence, leverage effect and fat tail behavior of real world data.



Graph 3.1.1: ARMA(1,1)FIARCHCh(2,1)Skewed-t.

Best estimator in test, Log-L=4241,82. Note that the series and the residuals are almost identical implying a good reproduction of characteristics, outliers are perfectly cached.



Graph 3.1.2: GARCH(1,1)Normal.

One of the worse estimators in comparison, Log-L=4168.74. Reproduction of the data was able to catch the important outliers but tends to stay closer around mean. Residuals graph is "thicker". Note the highest two points around data 1200 (November-December 2000). Conditional variance is higher where medium size residuals of around ± 0.1 cluster than where single big residual of around -0.2. This is the opposite in above graph. GARCH (1, 1) Normal has slower responses.

Table 3.1.5: Test Statistics Of Each Model

		Information Criterium (minimize)			
Model	Log-L	Akaike	Schwarz	Shibata	Hannan-Quinn
FIAparchCh11Skt	4237.81	-4.230929	-4.20571	-4.230969	-4.221670
FIAparchBBM11Skt	4237.78	-4.230895	-4.20568	-4.230935	-4.221636
FIAparchBBM11St-t	4236.83	-4.230948	-4.20854	-4.230980	-4.222719
FIAparchCh11St-t	4236.47	-4.230588	-4.20818	-4.230620	-4.222358
HYGarch11Skt	4234.68	-4.228795	-4.20638	-4.228827	-4.220565
HYGarch11St-t	4234.04	-4.229157	-4.20955	-4.229182	-4.221956
FIgarchCh11Skt	4233.97	-4.229089	-4.20948	-4.229114	-4.221888
FIGarchBBM11Skt	4233.81	-4.228926	-4.20931	-4.228950	-4.221725
FIgarchCh11St-t	4233.39	-4.229502	-4.21269	-4.229520	-4.223330
FIgarchBBM11St-t	4233.25	-4.229369	-4.21256	-4.229387	-4.223197
Aparch11Skt	4228.85	-4.222966	-4.20055	-4.222998	-4.214736

Table 3.1.5 devamı

Gjr11Skt	4228.82	-4.223935	-4.20432	-4.223959	-4.216734
Aparch11St-t	4227.98	-4.223096	-4.20349	-4.223120	-4.215895
Gjr11St-t	4227.97	-4.224081	-4.20727	-4.224099	-4.217909
FIAparchCh11GED	4226.78	-4.220889	-4.19848	-4.220920	-4.212659
FIAparchBBM11GED	4226.64	-4.220747	4.19834	-4.220779	-4.212518
Garch11Skt	4225.11	-4.221221	-4.20441	-4.22124	-4.215049
HYGarch11GED	4224.61	-4.219716	-4.20011	-4.219741	-4.212515
Garch11St-t	4224.59	-4.221704	-4.20770	-4.22172	-4.216561
FIgarchCh11GED	4224.27	-4.220382	-4.20357	-4.220400	-4.214210
FIgarchBBM11GED	4224.05	-4.220160	-4.20335	-4.220178	-4.213987
Igarch11Skt	4222.01	-4.219117	-4.20511	-4.219129	-4.213973
Igarch11St-t	4221.30	-4.219411	-4.20821	-4.219419	-4.215296
Gjr11GED	4218.37	-4.214472	-4.19766	-4.214490	-4.208300
Aparch11Ged	4218.36	-4.213471	-4.19386	-4.213496	-4.206270
Garch11GED	4216.12	-4.213226	-4.19922	-4.21324	-4.208082
Igarch11GED	4212.15	-4.210254	-4.19905	-4.210262	-4.206139
EGarch11St-t	4205.20	-4.200303	-4.18069	-4.20033	-4.193102
FIEgarch11Skt	4203.91	-4.197007	-4.17179	-4.197047	-4.187749
FIEgarch11GED	4192.53	-4.186621	-4.16421	-4.186653	-4.178392
FIAparchCh11N	4182.80	-4.177891	-4.15828	-4.177916	-4.170690
FIAparchBBM11N	4181.37	-4.176462	-4.15685	-4.176486	-4.169261
FIgarchCh11N	4179.29	-4.176376	-4.16237	-4.176389	-4.171233
HYGarch11N	4179.07	-4.175160	-4.15835	-4.175178	-4.168988
FIgarchBBM11N	4178.99	-4.176076	-4.16207	-4.176088	-4.170932
EGarch11Skt	4172.21	-4.166289	-4.14388	-4.16632	-4.158059
Aparch11N	4170.24	-4.166320	-4.14951	-4.166338	-4.160148
Gjr11N	4170.24	-4.167319	-4.15331	-4.167331	-4.162175
Garch11N	4168.74	-4.166826	-4.15562	-4.16683	-4.162711
Igarch11N	4163.80	-4.162879	-4.15447	-4.162883	-4.159793
EGarch11GED	4154.92	-4.149993	-4.13038	-4.15002	-4.142792
FIEgarch11St-t	4122.17	-4.116231	-4.09382	-4.116263	-4.108002
FIEgarch11N	4120.69	-4.115746	-4.09614	-4.115770	-4.108545
EGarch11N	4096.69	-4.092739	-4.07593	-4.09276	-4.086567

3.2 Forecasting Results

As we expected, the best models for estimation are not necessarily the best ones for forecasting. The same thing is also true for the distributions. The specification of the model has a more clear and predictable effect on Mincer Zarnowitz regression R^2. As explained by Laurent and Peters (2002) the Mincer-Zarnowitz regression has been largely used to evaluate forecasts in the conditional mean. For the conditional variance, it is computed by regressing the forecasted variances on the actual variances.

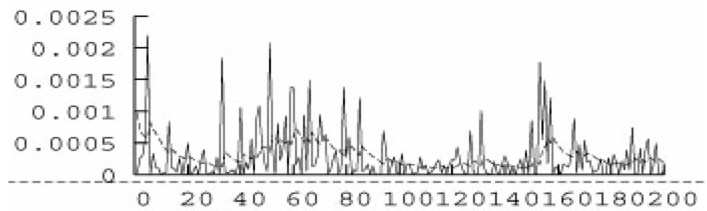
$$\sigma_t^2 = \alpha + \beta\hat{\sigma}_t^2 + \upsilon_t \tag{34}$$

The other criteria are minimizing errors and lead to difficult interpretation and inconsistent rankings. Like the maximum likelihood in estimation, R^2 is in general more consistent with the aggregated ranking results.

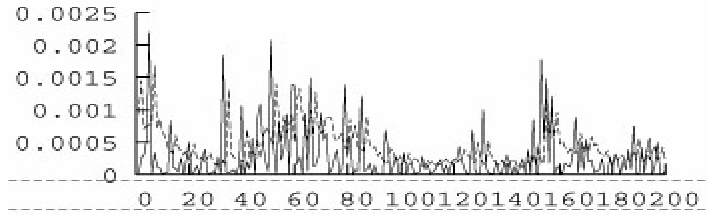
Table 3.2.1: 20 Best Forecasting Models Based On R^2

	Model Structure				
Model	Mean Equation	Variance Equation	Distribution	Log-L	R^2
Igarch11Skt	ARMA (0,0)	Igarch (1,1)	Skewed-t	4222.01	0.021975
Igarch11N	ARMA (0,0)	Igarch (1,1)	Normal	4163.80	0.021846
Igarch11GED	ARMA (0,0)	Igarch (1,1)	GED	4212.15	0.021611
Aparch33Skt	ARMA (0,0)	Aparch (3,3)	Skewed-t	4236.68	0.021544
Igarch11St -t	ARMA (0,0)	Igarch (1,1)	Student-t	4221.30	0.021478
Gjr33Skt	ARMA (0,0)	GJR (3,3)	Skewed-t	4236.68	0.021404
Ar1Igarch11Skt	ARMA (1,0)	Igarch (1,1)	Skewed-t	4223.65	0.020980
ARMA11Igarch11GED	ARMA (1,1)	Igarch (1,1)	GED	4214.04	0.020875
ARMA11Igarch11Skt	ARMA (1,1)	Igarch (1,1)	Skewed-t	4225.31	0.020874
Aparch11Skt	ARMA (0,0)	Aparch (1,1)	Skewed-t	4228.85	0.020827
Gjr11Skt	ARMA (0,0)	GJR (1,1)	Skewed-t	4228.82	0.020781
ARMA11GJR33Skt	ARMA(1,1)	GJR (3,3)	Skewed-t	4239.43	0.020429
Aparch11St-t	ARMA (0,0)	Aparch (1,1)	Student-t	4227.98	0.020152
Gjr11St-t	ARMA (0,0)	GJR (1,1)	Student-t	4227.97	0.020110
Aparch11N	ARMA (0,0)	Aparch (1,1)	Normal	4170.24	0.020053
Gjr11N	ARMA (0,0)	GJR (1,1)	Normal	4170.24	0.020052
Gjr11GED	ARMA (0,0)	GJR (1,1)	GED	4218.37	0.019908
Aparch11Ged	ARMA (0,0)	Aparch (1,1)	GED	4218.36	0.019904
AR1Aparch11Skt	ARMA (1,0)	Aparch (1,1)	Skewed-t	4230.44	0.019801
Garch11Skt	ARMA (0,0)	GARCH (1,1)	Skewed-t	4225.11	0.019484

IGARCH performance is worth noting. Riskmetrics process of J.P. Morgan is also a kind of IGARCH. For details of the model see Mina and Xiao (2001).

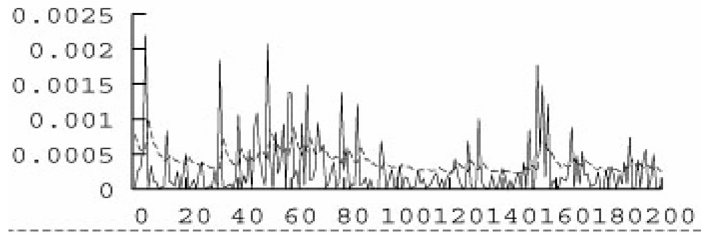


Graph 3.2.1: IGARCH(1,1) Skt is one of the integrated models that proved to be a good forecast model based on R^2. MSE tells it is the third worse.



Graph 3.2.2: EGARCH(1,1) GED converged only weakly after 208 BFGS iterations in 52 seconds. However it reached a record level R^2 and MSE.

Graphically it seems to make a fairly good conditional variance forecast of absolute returns. A closer look predicts that the peak points of forecasts follow actual data with a small delay. Based on MAE and RMSE this model is the second worse.



Graph 3.2.3: ARMA(1,1)FIAPARCH(2,d,1) Skt was the best estimator according to Log-L, an average forecaster based on R^2. RMSE ranks it to the bottom.

GED and skewed-t distributions performed well in predictions but it is not possible to favor any distribution. While RMSE and MAE ranked the t-distributions better, MSE favored GED and puts t distributions to the bottom. R^2 made no implications on the distribution. Normal distributions give consistently moderate results.

Table 3.2.2: First 20 And Last Ten Models With Constant Mean İn Forecasting Performances According To Different Criteria

	R^2	RMSE	R^2 Rank	MSE	R^2 Rank	MAE	R^2	
5	EGarch11GED	Garch11Skt	14	Gjr11GED	12	Garch11GED	17	5
	Igarch11Skt	FIAparchBBM11Skt	19	Aparch11Ged	13	FIEgarch11Skt	40	
	Igarch11N	FIAparchCh11Skt	18	EGarch11GED	1	Garch11Skt	14	
	Igarch11GED	FIGarchBBM11Skt	25	FIEgarch11N	38	FIAparchBBM11Skt	19	
	Igarch11St-t	Igarch11Skt	2	Aparch11N	10	FIAparchCh11Skt	18	
10	Aparch11Skt	HYGarch11Skt	21	Gjr11N	11	FIGarchBBM11Skt	25	10
	Gjr11Skt	FIgarchCh11Skt	22	FIAparchBBM11GED	20	Igarch11Skt	2	
	Aparch11St-t	Aparch11Skt	6	FIAparchCh11GED	36	HYGarch11Skt	21	
	Gjr11St-t	Gjr11Skt	7	FIEgarch11GED	39	FIgarchCh11Skt	22	
	Aparch11N	HYGarch11St-t	32	Garch11GED	17	FIAparchBBM11St-t	23	
15	Gjr11N	FIgarchCh11St-t	33	FIgarchCh11GED	28	FIAparchCh11St-t	37	15
	Gjr11GED	FIEgarch11Skt	40	FIAparchCh11N	35	Igarch11St-t	5	
	Aparch11Ged	Igarch11N	3	Igarch11GED	4	Garch11St-t	16	
	Garch11Skt	FIAparchBBM11St-t	23	FIAparchBBM11N	24	Aparch11Skt	6	
	Garch11N	FIAparchCh11St-t	37	HYGarch11GED	26	Gjr11Skt	7	
20	Garch11St-t	Igarch11St-t	5	FI-garchBBM11GED	31	HYGarch11St-t	32	20
	Garch11GED	Garch11St-t	16	Aparch11St-t	8	FIgarchCh11St-t	33	
	FIAparchCh11Skt	FIgarchBBM11St-t	34	Gjr11St-t	9	FIgarchBBM11St-t	34	
	FIAparchBBM11Skt	Gjr11GED	12	Garch11N	15	FIgarchCh11GED	28	
	FIAparchBBM11GED	Aparch11Ged	13	FIgarchBBM11N	29	FIAparchCh11N	35	
	…………							
10	FIgarchBBM11GE	HYGarch11GED	26	FIgarchCh11St-t	33	FIAparchCh11GED	36	10
	HYGarch11St-t	FIgarchBBM11GED	31	FIgarchBBM11St-t	34	FIEgarch11GED	39	
	FIgarchCh11St-t	Aparch11St-t	8	FIEgarch11Skt	40	Igarch11GED	4	
	FIgarchBBM11St-t	Gjr11St-t	9	Garch11Skt	14	FIAparchBBM11N	24	
	FIAparchCh11N	Garch11N	15	FIAparchBBM11Sk	19	HYGarch11GED	26	
5	FIAparchCh11GED	FIgarchBBM11N	29	FIAparchCh11Skt	18	FI-garchBBM11GED	31	5
	FIAparchCh11St-t	HYGarch11N	30	FIGarchBBM11Skt	25	Gjr11GED	12	
	FIEgarch11N	FIgarchCh11N	27	Igarch11Skt	2	Aparch11Ged	13	
	FIEgarch11GED	EGarch11GED	1	HYGarch11Skt	21	EGarch11GED	1	
	FIEgarch11Skt	FIEgarch11N	38	FIgarchCh11Skt	22	FIEgarch11N	38	

While R^2 ranks according to model specification, minimum error criteria seem to give more importance to distributions. FI (1) models perform poor forecasts based on R^2 whereas other criteria do not allow for a conclusion.

It is also hard to draw conclusions from the model specification. Remarkable are the forecasting performances of GJR, IGARCH and GARCH. R^2 gave the worst performances with FI (1) models while the best performers were I (1) and non-integrated ones. Therefore we can conclude that either a complete integration or no integration is preferred to a fractional integration. In general one obtains better R^2 results the simpler a model is specified. Increased parameters through modifications in mean or higher orders provide poor R^2 results. Especially the order (2, 2) consistently outputs very poor R^2. The order (3, 3) can be either a good performer or a bad choice, but it is worth trying. The results of FI processes of BMM and Chung are again very similar. According to the evaluation criteria they are either among the first or among the very last.

Table 3.2.3: Minimum Sum Of Rankings For Different IGARCH Specifications

	R^2	RMSE	MSE	MAE	TOTAL
Igarch11N	2	2	2	3	9
Igarch11GED	3	1	1	4	9
Igarch11Skt	1	4	5	1	11
Igarch11St-t	4	2	3	2	11
Igarch33Skt	10	3	4	1	18
Igarch22Skt	8	4	6	1	19
Ar1Igarch11Skt	5	5	9	3	22
ARMA11Igarch11GED	6	4	7	5	22
AR1Igarch22Skt	9	5	8	2	24
ARMA11Igarch11Skt	7	6	10	6	29

R^2 rankings are consistent with aggregate rankings. In general simpler models with less parameter perform better forecasts.

Table 3.2.4: Minimum Sum Of Rankings For Different GARCH Specifications

	R^2	RMSE	MSE	MAE	TOTAL
Garch11GED	4	1	1	1	7
Garch11N	2	1	2	4	9
Garch11St-t	3	2	3	3	11
Garch11Skt	1	4	6	2	13
AR1Garch11N	6	3	4	5	18
AR1Garch11Skt	5	4	8	4	21
AR1Garch-m11Skt	7	3	5	6	21
Garch22	9	4	7	2	22
ARMA22Garch-m33Skt	8	3	6	7	24
AR1Garch22Skt	10	4	8	3	25

RMSE, MSE and MAE give very different rankings in cross comparison. However in general mean specifications restrict the flexibility of all models and result in a general trend and can not capture the outliers.

Simple GARCH(1,1) performs generally well according to all criteria. GARCH estimation outputs the sum of all coefficients very close to 1. This explains its forecasting success close to IGARCH. The size of the sample is a crucial factor affecting the forecasting performance. Therefore we believe that most models would behave differently with different sample sizes which could be the topic of a separate research. Test statistics for forecast evaluation measures of some models are given in the following Table.

Table 3.2.5: Test Statistics Of Some Models İn Forecasting Performances According To Different Criteria

	Forecast Evaluation Measures						
Model	**R^2**	**MSE(M)**	**MSE(V)**	**MAE(M)**	**MAE(V)**	**RMSE (M)**	**RMSE (V)**
ARMA22Garch-m33Skt	0,014413	0.0002767	1.705E-07	0.01307	0.000271	0.01663	0.000413
AR1Garch22Skt	0.008062	0.0002770	1.708E-07	0.01299	0.000284	0.01664	0.000413
Garch22Skt	0.008096	0.0002768	1.708E-07	0.01298	0.000283	0.01664	0.000413
AR1Garch-m11Skt	0.017135	0.0002766	1.689E-07	0.01306	0.000271	0.01663	0.000411
AR1Garch11Skt	0.018712	0.0002770	1.680E-07	0.01300	0.000271	0.01664	0.000410
Garch11Skt	0.019483	0.0002767	1.678E-07	0.01298	0.000270	0.01664	0.000410
Garch11St-t	0.018915	0.0002763	1.682E-07	0.01299	0.000270	0.01662	0.000410
EGarch11St-t	0.008337	0.0002758	1.900E-07	0.01301	0.000267	0.01661	0.000436
ARMA11Egarch22Skt	0.020393	0.0002801	2.616E-07	0.01305	0.000391	0.01674	0.000512
AR1Egarch11Skt	0.028130	0.0002779	2.713E-07	0.01299	0.000386	0.01667	0.000521
EGarch11Skt	0.028791	0.0002780	2.743E-07	0.01297	0.000388	0.01667	0.000524
ARMA11Egarch22GED	0.023115	0.0002769	2.227E-07	0.01307	0.000357	0.01664	0.000472
EGarch11GED	0.028731	0.0002757	2.290E-07	0.01304	0.000354	0.01660	0.000479
EGarch11N	0.028718	0.0002757	2.143E-07	0.01306	0.000349	0.01660	0.000463
ARMA11GJR33Skt	0.020429	0.0002777	1.690E-07	0.01304	0.000272	0.01667	0.000411
AR1Gjr22Skt	0.008727	0.0002767	1.729E-07	0.01299	0.000288	0.01663	0.000416
Gjr33Skt	0.021404	0.0002763	1.692E-07	0.01299	0.000271	0.01662	0.000411
Gjr11Skt	0.020781	0.0002764	1.689E-07	0.01299	0.000273	0.01663	0.000411
Gjr11St-t	0.020110	0.0002760	1.692E-07	0.01300	0.000273	0.01661	0.000411
AR1Gjr11GED	0.019125	0.0002759	1.688E-07	0.01302	0.000270	0.01661	0.000411
Gjr11GED	0.019908	0.0002757	1.685E-07	0.01302	0.000270	0.01661	0.000411
Gjr33N	0.018225	0.0002759	1.697E-07	0.01301	0.000271	0.01661	0.000412
ARMA11GJR22N	0.008162	0.0002776	1.706E-07	0.01309	0.000281	0.01666	0.000413
Gjr11N	0.020052	0.0002758	1.678E-07	0.01301	0.000269	0.01661	0.000410
ARMA11Aparch33Skt	0.008243	0.0002780	1.731E-07	0.01304	0.000289	0.01667	0.000416
AR1Aparch22Skt	0.008754	0.0002766	1.730E-07	0.01299	0.000288	0.01663	0.000416
Aparch22Skt	0.008964	0.0002765	1.727E-07	0.01298	0.000287	0.01663	0.000416
Aparch33Skt	0.021544	0.0002763	1.691E-07	0.01299	0.000271	0.01662	0.000411
AR1Aparch11Skt	0.019801	0.0002766	1.692E-07	0.01300	0.000274	0.01663	0.000411

Table 3.2.5. Devamı

Aparch11Skt	0.020827	0.0002764	1.688E-07	0.01299	0.000272	0.01663	0.000411
Aparch11St-t	0.020152	0.0002760	1.691E-07	0.01300	0.000272	0.01661	0.000411
ARMA11Aparch-m11GED	0.017656	0.0002786	1.693E-07	0.01325	0.000267	0.01669	0.000411
Aparch11Ged	0.019904	0.0002757	1.685E-07	0.01302	0.000270	0.01661	0.000411
Aparch11N	0.020053	0.0002758	1.678E-07	0.01301	0.000269	0.01661	0.000410
Igarch33Skt	0.007316	0.0002767	1.773E-07	0.01298	0.000304	0.01663	0.000421
AR1Igarch22Skt	0.008869	0.0002771	1.759E-07	0.01299	0.000303	0.01665	0.000419
Igarch22Skt	0.008889	0.0002769	1.760E-07	0.01298	0.000303	0.01664	0.000420
ARMA11Igarch11Skt	0.020874	0.0002789	1.714E-07	0.01305	0.000294	0.01670	0.000400
Ar1Igarch11Skt	0.020980	0.0002772	1.714E-07	0.01300	0.000293	0.01665	0.000414
Igarch11Skt	0.021975	0.0002769	1.710E-07	0.01298	0.000293	0.01664	0.000414
Igarch11St-t	0.021478	0.0002763	1.715E-07	0.01299	0.000293	0.01662	0.000414
ARMA11Igarch11GED	0.020875	0.0002770	1.709E-07	0.01305	0.000293	0.01664	0.000413
Igarch11GED	0.021611	0.0002759	1.706E-07	0.01301	0.000292	0.01661	0.000413
Igarch11N	0.021846	0.0002761	1.701E-07	0.01300	0.000292	0.01662	0.000412
AR1FIgarchBBM22St-t	0.012166	0.0002765	1.716E-07	0.01299	0.000303	0.01663	0.000414
FIGarchBBM11Skt	0.012040	0.0002768	1.711E-07	0.01298	0.000302	0.01664	0.000414
FIgarchBBM11St-t	0.011605	0.0002764	1.713E-07	0.01299	0.000302	0.01662	0.000414
FIgarchBBM11GED	0.011802	0.0002759	1.707E-07	0.01301	0.000301	0.01661	0.000413
FIgarchBBM11N	0.011863	0.000276	1.708E-07	0.01300	0.000302	0.01661	0.000413

4. Conclusion

Most linear time series models for prediction of returns descend from the AutoRegressiv Moving Average (ARMA) and Generalized Autoregressiv Conditional Heteroskedastic (GARCH) models. Both concepts are useful in volatility modeling, but less useful in return prediction.

Scientific prediction involves the spotting of past patterns or regularities and testing them on recent observations. The data used to spot the patterns can therefore be called the training data. Parametric models like GARCH make use of the training data to modify the parameters in such a way that it fits best to the data. As a consequence well structured models are able to model the data almost precisely. However in the attempt to predict the future values with the same model one actually assumes that the future results will follow the same characteristics, same patterns. It is also assumed that the reactions to factors not included in the model are similar in both the past and the future. This is the reason why GARCH models as parametric specifications operate best under relatively stable market conditions. Although GARCH is explicitly designed to model time-varying conditional variances, GARCH models can fail to predict highly irregular phenomena, including wild market fluctuations (e.g., crashes and subsequent rebounds), and other highly unanticipated events that can lead to significant structural change. The choice of the optimum sample size, the window size for p and q are still highly an art based on experience.

In this study, in order to the estimation and forecasting performances of different GARCH processes, Ox 3.40 and its G@RCH 3.0 packages dedicated to GARCH models and many of its extensions by Laurent and Peters is used for analyzing the models. With open-source code are able to add or modify specifications, processes or graphics in the future.

The $(p,q) = (1,1)$ variant of all models are systematically tested with four different distributions, namely, Gaussian Normal, Student-t, Generalized Error Distribution (GED) and Skewed-t distribution. 94 models of higher order are unsystematically experimented to examine the effects on estimation and forecasting performances with an emphasis on Maximum Likelihood.

Estimation results are evaluated on the basis of ML, Akaike, Schwarz, Shibata and Hannan-Quinn values, whereas forecasting results are ranked according to Mincer Zarnowitz regression R^2, Root Mean Square Error, Mean Square Error and Mean Absolute Error value criteria. In comparing the estimation powers of models, we restrict our comments to student and skewed-t distributions. For optimizing maximum likelihood, skewed-t performs better than student-t for all models. On the other hand if we evaluate according to the other four criteria, by which more complicated models are penalized for the inclusion of additional parameters, skewed-t looses its apparent advantage, because it requires an additional skewness parameter. Especially Hannan-Quinn test seems to judge according to the distribution rather than the model spesification and prefer student-t. We found that the choice of models is at least as important as the choice of distributions, because best performing models combined with both distributions found place in the front ranks, mostly successively.

Log Likelihood results are consistent with aggregate results. Normal distribution estimates worse for all models. Higher orders alone improve results more than ARMA specifications alone. Together they improve more but the marginal benefit decreases. As a result we can conclude that, based on the maximum likelihood, the more complex a model is the better it fits to the data.

Combined rankings allow for the conclusion that maximum likelihood is a consistent evaluation criterion. While Akaike, Shibata, Schwarz and Hannan-Quinn may result in complete different rankings, the aggregate results are consistent with that of the maximum likelihood.

If skewed-t or student-t distributions are used, it is always possible to increase the likelihood by adding more tailored processes, implying a better fit to the data on the basis of numbers. GARCH models in general succeed in reproducing volatility clustering, persistence, leverage effect and fat tail behavior of real world data.

For forecasting, the best models for estimation are not necessarily the best ones. The same thing is also true for the distributions. The specification of the model has a more clear and predictable effect on Mincer-Zarnowitz regression R^2. As explained by Laurent and Peters (2002) the Mincer-Zarnowitz regression has been largely used to evaluate forecasts in the conditional mean. For the conditional variance, it is computed by regressing the forecasted variances on the actual variances.

GED and skewed-t distributions performed well in predictions but it is not possible to favor any distribution. While RMSE and MAE ranked the t-distributions better, MSE favored GED and puts t distributions to the bottom. R^2 made no implications on the distribution. Normal distributions give consistently moderate results. In general simpler models with less parameter perform better forecasts. RMSE, MSE and MAE give very different rankings in cross comparison. However in general mean specifications restrict the flexibility of all models and result in a general trend and can not capture the outliers.

The size of the sample is a crucial factor affecting the forecasting performance. Therefore we believe that most models would behave differently with different sample sizes which could be the topic of a separate research.

Volatilite Değerleme ve Tahmini Için ARCH ve GARCH Modellerinin Kullanımı

Özet: Bu çalışma, 9 yıllık günlük verilere dayanarak IMKB 100 endeksinin volatilitesini değerlendirmek ve tahmin etmek için, her biri dört ayrı dağılımla denenen, ARMA özellikleri eklenebilen 11 değişik ARCH modelinin performansını sunmaktadır. Elde edilen sonuçlara göre, aynı dağılım kullanılırsa, kısmi entegre edilmiş asimetrik modeller bu özelliğe sahip olmayan orjinal versiyonlarından daha iyi volatilite değerlemesi yapabilmektedir. Çarpık-t ve Student-t dağılımlarının kullanılması modelin veriye daha uyumlu olmasını sağlamaktadır. Sonuç olarak, belirli bir model veya dağılımın kullanılmasının volatilite tahmininde açık bir iyileşmeye yol açmadığı gözlenmiştir.

Anahtar Kelimeler: GARCH; EGARCH; GJR; APARCH; IGARCH; FIGARCH; FIAPARCH; FIEGARCH; HYGARCH; ARMA; GED; Skewed-t; Ox; G@RCH

References

Baillie, R., T. Bollerslev, and H. Mikkelsen (1996): "Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 74, 3–30., December 17, 2004. http://www.sciencedirect.com

Bhardwaj, G. and Swanson, N. R. (2004): "An Empirical Investigation of the Usefulness of ARFIMA Models for Predicting Macroeconomic and Financial Time Series" *Special issue of the Journal of Econometrics on "Empirical Methods in Macroeconomics and Finance".*, November 15, 2004. ftp://snde.rutgers.edu/Rutgers/wp/2004-22.pdf

Bollerslev, T. (1986): "Generalized autoregressive condtional heteroskedasticity," *Journal of Econometrics*, 31, 307–327., September 20, 2004. http://www.sciencedirect.com

Bollerslev, T., and H. O. Mikkelsen (1996): "Modeling and Pricing Long-Memory in Stock Market Volatility," *Journal of Econometrics*, 73, 151–184., December 7, 2004. http://www.sciencedirect.com

Chung, C. F. (1999): "Estimating the fractionnally intergrated GARCH model," National Taiwan University working paper, December 17, 2004. http://gate.sinica.edu.tw/~metrics/Pdf_Papers/Figarch.pdf

Davidson, J. (2001): "Moment and Memory Properties of Linear Conditional Heteroscedasticity Models," Manuscript, Cardiff University, January 4, 2005. http://www.ex.ac.uk/~jehd201/hygarch4.pdf

Ding, Z., C. W. J. Granger, and R. F. Engle (1993): "A Long Memory Property of Stock Market Returns and a New Model," *Journal of Empirical Finance*, 1, 83–106., September 20, 2004. http://www.sciencedirect.com

Doornik, J. A. (1999): An Object Oriented Matrix Programming Language. Timberlake Consultant Ltd., third edn., *Ox Documentation*.

Doornik, J., and M. Ooms (1999): "A Package for Estimating, Forecasting and Simulating ARFIMA Models: Arfima Package 1.0 for Ox," Nuffield College (Oxford) discussion paper, January 4, 2005. http://www.doornik.com/download/arfima.pdf

Engle, R. (1982): "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation," *Econometrica*, 50, 987–1007., December 7, 2004. http://www.sciencedirect.com

Engle, R., and T. Bollerslev (1986): "Modeling the Persistence of Conditional Variances," *Econometric Reviews*, 5, 1–50., December 17, 2004. http://www.sciencedirect.com

Glosten, L., R. Jagannathan, and D. Runkle (1993): "On the relation between expected return on stocks," *Journal of Finance*, 48, 1779–1801., November 15, 2004. http://www.sciencedirect.com

Holton, G. (1996): "Contingency Analysis" family of websites including a riskglossary, November 10, 2004 http://www.contingencyanalysis.com, www.riskglossary.com

Laurent, S., and J. P. Peters (2002): "A tutorial for G@RCH 2.3: a complete Ox Package for Estimating and Forecasting ARCH Models," *G@RCH 3.0 documentation*.

Lambert, P., and S. Laurent (2001): "Modelling Financial Time Series Using GARCH-Type Models and a Skewed Student Density," Mimeo, Universite de Liege., November 10, 2004. http://www.core.ucl.ac.be/~laurent/pdf/Lambert-Laurent.pdf

McKenzie, M. and Mitchell, H. (2001): "Generalised Asymetric Power ARCH Modelling of Exchange Rate Volatility" Royal Melbourne Institute of Technology discussion paper, January 4, 2005. http://mams.rmit.edu.au/mmt5alsrzfd2.pdf

Mina, J. and Xiao, J..Y. (2001): "Return to RiskMetrics: The Evolution of a Standard" Copyright © 2001 RiskMetrics Group. Update and restatement of the mathematical models in the 1996 *RiskMetrics Technical Document*, November 2, 2004. http://www.riskmetrics.com

Nelson, D. (1991): "Conditional heteroskedasticity in asset returns: a new approach," *Econometrica*, 59, 349–370., December 7, 2004. http://www.sciencedirect.com

Peters, J.P. (2001) "Estimating and Forecasting Volatility of Stock Indices Using Asymmetric GARCH Models and Skewed Student-t Densities," Working Paper, École d'Administration des Affaires, University of Liège, Belgium, January 3, 2005. http://www.panagora.com/2001crowell/2001cp_50.pdf

Tse, Y. (1998): "The Conditional Heteroscedasticity of the Yen-Dollar Exchange Rate," *Journal of Applied Econometrics*, 193, 49–55., November 10, 2004. http://qed.econ.queensu.ca:80/jae/1998-v13.1/ by http://ideas.repec.org

Zakoian, J. M. (1994): "Threshold Heteroscedastic Models," *Journal of Economic Dynamics and Control*, 18, 931–955., November 10, 2004. http://papers.ssrn.com

Zemke, S. (2003): "Data Mining for Prediction. Financial Series Case" The Royal Institute of Technology Department of Computer and Systems Sciences Doctoral Thesis, December 4, 2004. http://szemke.math.univ.gda.pl/zemke2003PhD.pdf