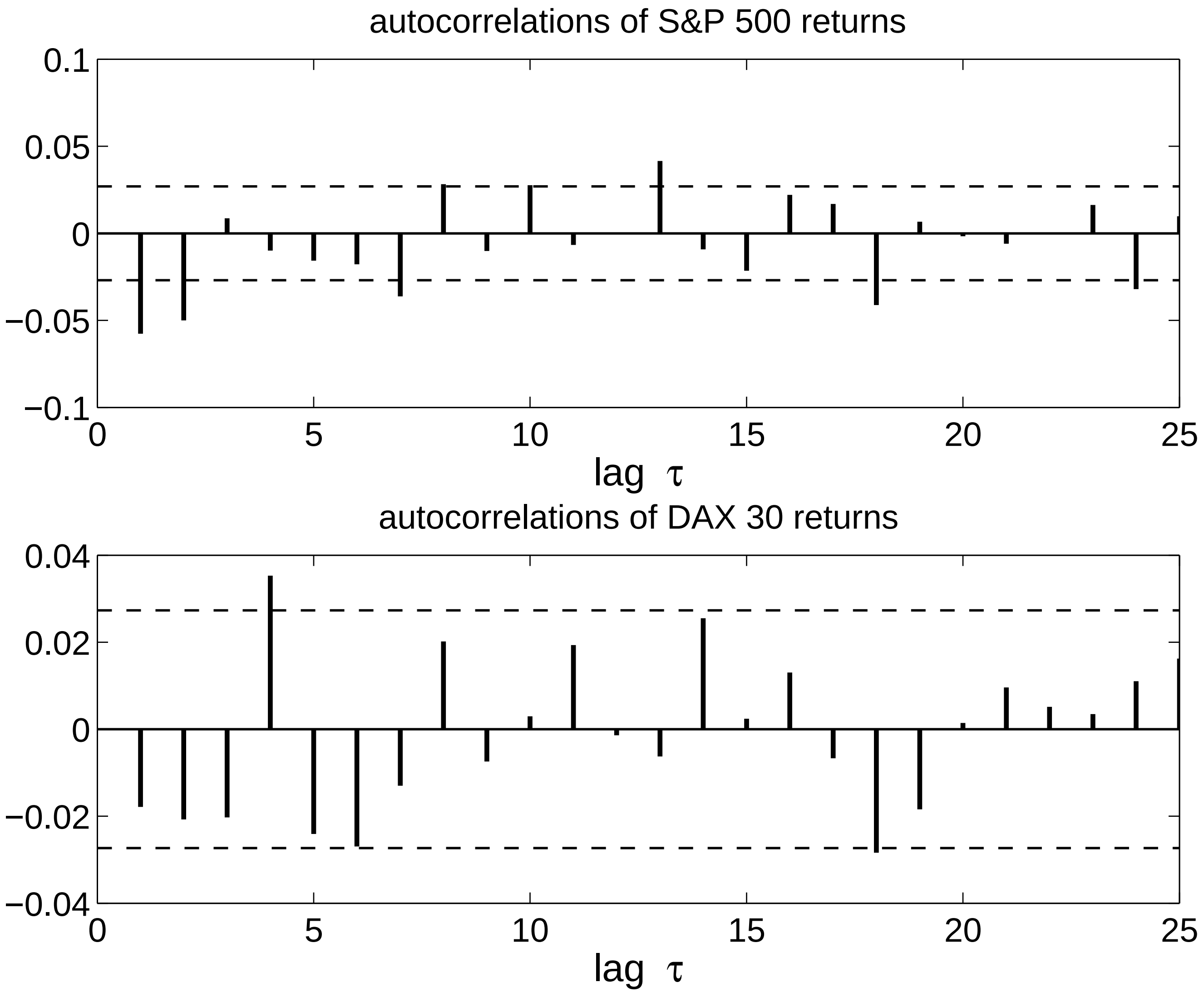
Financial Data Analysis

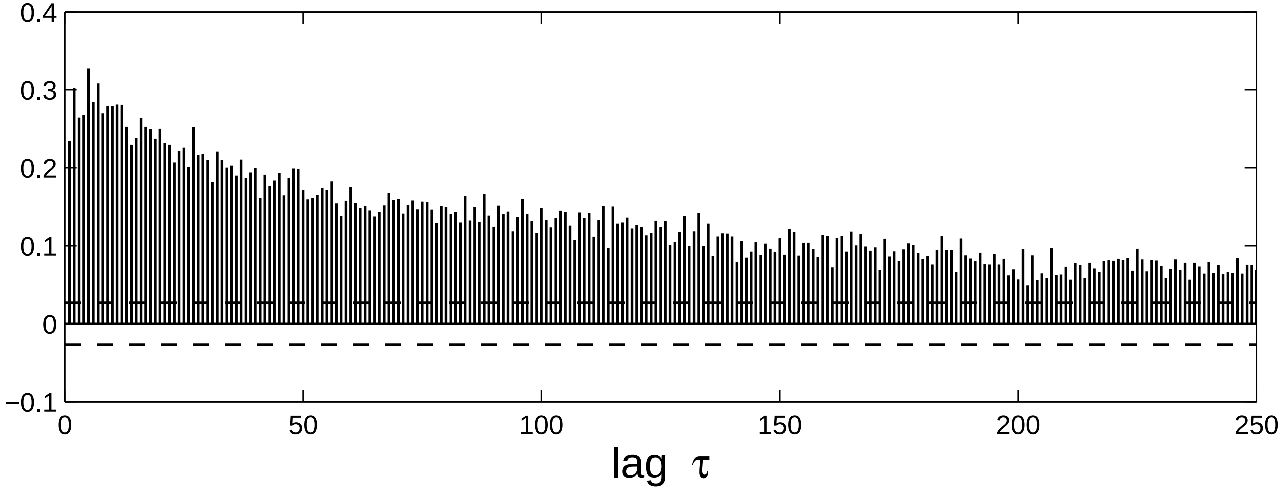
GARCH Models

July 12, 2011

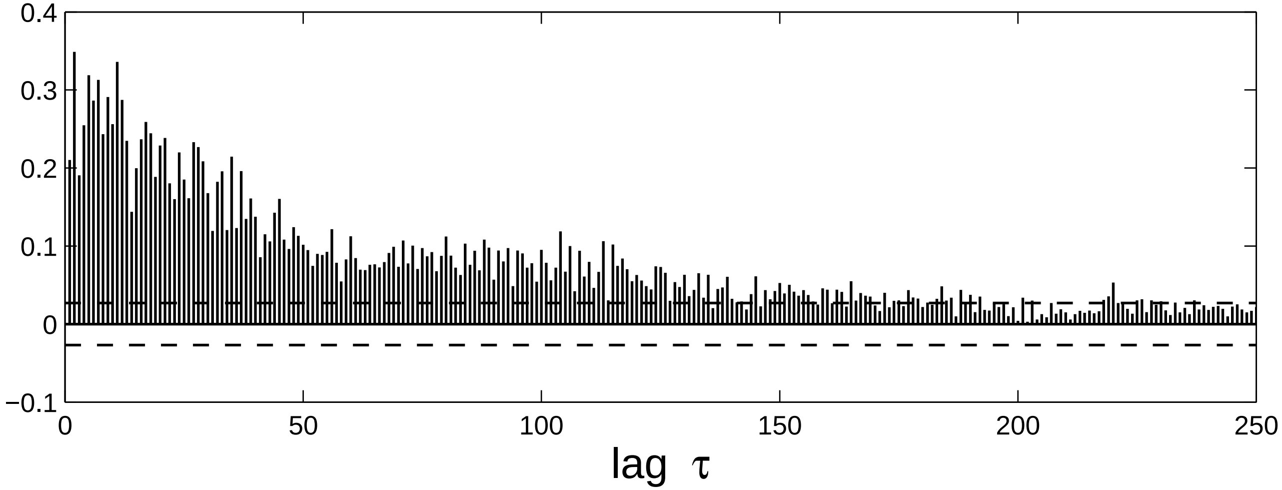


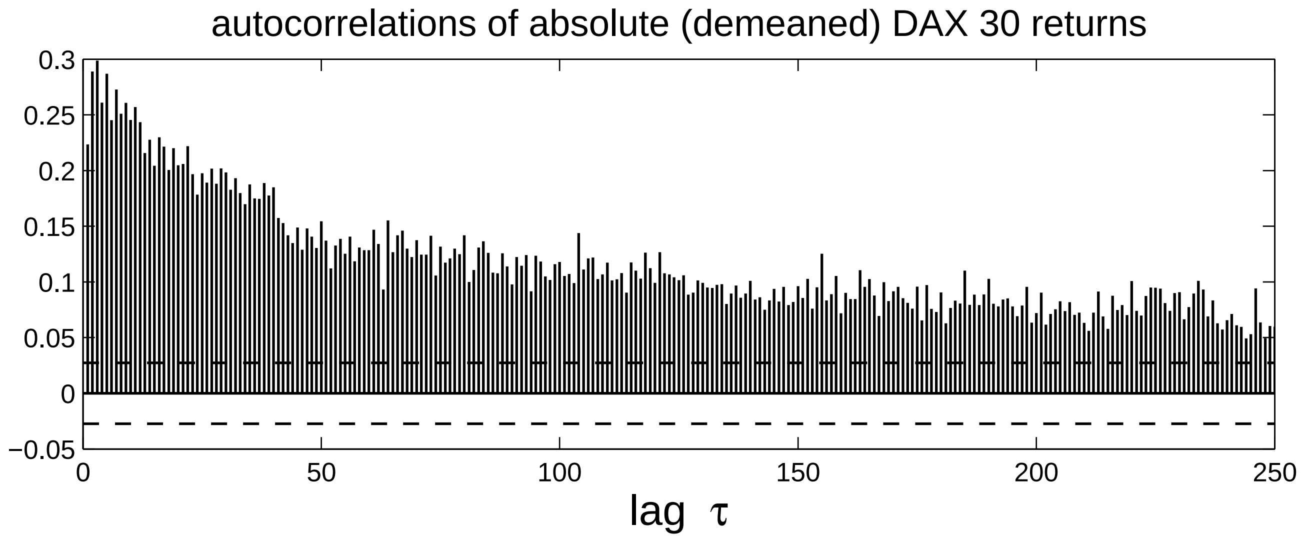
DAX: 1990–2009; S&P500: 1990–2010

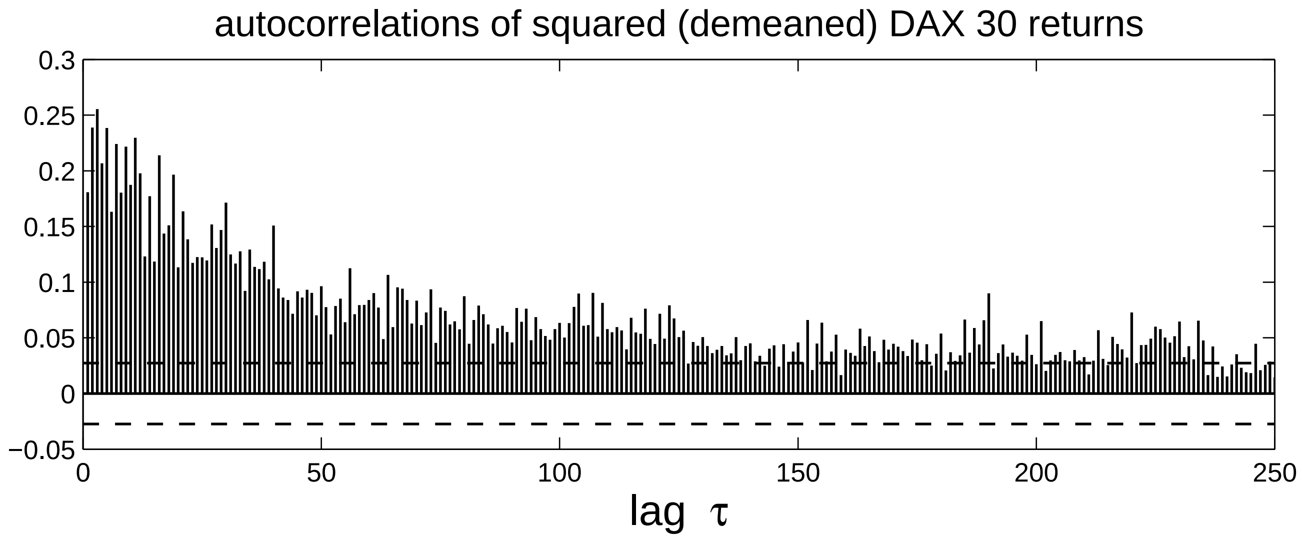
autocorrelations of absolute (demeaned) S&P 500 returns

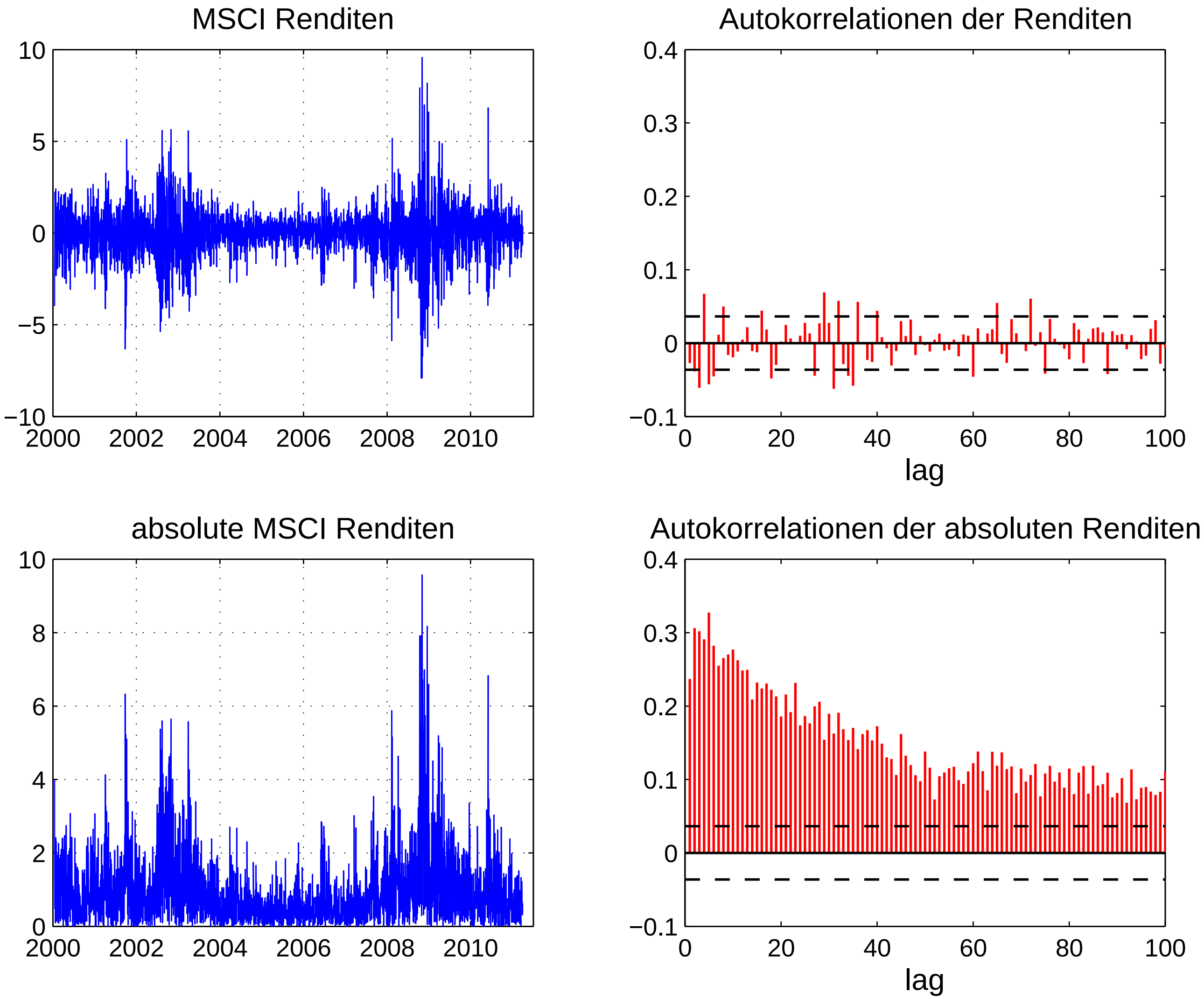


autocorrelations of squared (demeaned) S&P 500 returns




autocorrelations of absolute (demeaned) DAX 30 returns
0.3
0.25
0.2
0.15
0.1
0.05
0
−0.05
0
50
100
150
200
250
lag τ


autocorrelations of squared (demeaned) DAX 30 returns
0.3
0.25
0.2
0.15
0.1
0.05
0
−0.05
0
50
100
150
200
250
lag τ



MSCI Europe, 2000–2011

Several Stylized Facts

- Returns usually show no or only little autocorrelation.

- Volatility appears to be autocorrelated (volatility clusters).

- Normality is rejected in favor of a leptokurtic (fat–tailed) distribution.

Volatility Modeling and the Stylized Facts

- Consider the following model for returns r_t,

$$\begin{aligned} r_t &= \mu_t + \epsilon_t \qquad (1) \\ \epsilon_t &= \eta_t \sigma_t, \quad \eta_t \stackrel{iid}{\sim} N(0,1), \end{aligned}$$

  where we assume that the *innovation sequence* η_t is independent of σ_t.

- μ_t in (1) is the conditional mean of r_t conditional on the information up to time $t-1$. This may, for example, be constant or described by a low–order ARMA process.

- We are interested in the error term described by the second line of (1).

- If σ_t^2 depends on information available at time $t-1$, then σ_t^2 is the *conditional variance* of ϵ_t (and thus also r_t).

- Denote the information available up to time t by I_t; I_t typically consists of the past history of the process, $\{\epsilon_s : s \leq t\}$.

- Then we can also write

$$\epsilon_t | I_{t-1} \sim \mathsf{N}(0, \sigma_t^2), \tag{2}$$

i.e., ϵ_t is *conditionally* normally distributed with variance σ_t^2.

- However, if the conditional variance is time–varying (which is the case we are interested in), the *unconditional* distribution of ϵ_t will *not* be normal.

- To illustrate, consider the marginal kurtosis of ϵ_t, assuming ϵ_t is stationary with finite fourth moment,

$$\begin{aligned} \text{kurtosis}(\epsilon_t) &= \frac{\mathsf{E}(\epsilon_t^4)}{\mathsf{E}^2(\epsilon_t^2)} = \frac{\mathsf{E}(\eta_t^4 \sigma_t^4)}{\mathsf{E}^2(\eta_t^2 \sigma_t^2)} = \frac{\mathsf{E}(\eta_t^4)\mathsf{E}(\sigma_t^4)}{\underbrace{\mathsf{E}^2(\eta_t^2)\mathsf{E}^2(\sigma_t^2)}_{\text{Independence of } \eta_t \text{ and } \sigma_t^2}} \\ &= \underbrace{\frac{\mathsf{E}(\eta_t^4)}{\mathsf{E}^2(\eta_t^2)}}_{=3} \frac{\mathsf{E}(\sigma_t^4)}{\mathsf{E}^2(\sigma_t^2)} > 3, \end{aligned} \tag{3}$$

since

$$\mathsf{E}(\sigma_t^4) > \mathsf{E}^2(\sigma_t^2) \quad (\mathsf{E}(X^2) > \mathsf{E}^2(X)). \tag{4}$$

- An interpretation of (3) results from noting that

$$\begin{aligned}\frac{\mathsf{E}(\sigma_t^4)}{\mathsf{E}^2(\sigma_t^2)} &= 1 + \frac{\mathsf{E}(\sigma_t^4) - \mathsf{E}^2(\sigma_t^2)}{\mathsf{E}^2(\sigma_t^2)} \\ &= 1 + \frac{\mathsf{Var}(\sigma_t^2)}{\mathsf{E}^2(\sigma_t^2)}.\end{aligned}$$

- Thus, for a given level of the *unconditional variance* $\mathsf{E}(\sigma_t^2) = \mathsf{E}(\epsilon_t^2)$, the kurtosis of the marginal distribution of ϵ_t is increasing in the variability of the conditional variance.

- If $\mathsf{Var}(\sigma_t^2)$ is large, then σ_t^2 will often be considerably smaller (larger) then $\mathsf{E}(\sigma_t^2)$, giving rise to high peaks (thick tails) of the marginal distribution, respectively.

- Thus, even with normal innovations (conditional normality), time–varying conditional volatility may account for at least part of the leptokurtosis observed in financial return series.

- A further property of the error process is uncorrelatedness,

$$\mathsf{E}(\epsilon_t \epsilon_{t-\tau}) = \mathsf{E}(\eta_t \eta_{t-\tau} \sigma_t \sigma_{t-\tau}) = \underbrace{\mathsf{E}(\eta_t)}_{=0} \mathsf{E}(\eta_{t-\tau} \sigma_t \sigma_{t-\tau}) = 0.$$

- Absolute values and squares can be correlated, however, depending on the specification for the conditional variance process $\{\sigma_t^2\}$.

- Thus, at least in principle, a process of the form (1) is capable of reproducing several of the properties typically detected in financial returns.

The ARCH Process

- Engle (1982) introduced the class of **a**uto**r**regressive **c**onditional **h**eteroskedastic (ARCH) models,[1] where (1) is specified as

$$
\begin{aligned}
r_t &= \mu_t + \epsilon_t \\
\epsilon_t &= \eta_t \sigma_t, \quad \eta_t \overset{iid}{\sim} N(0,1), \\
\sigma_t^2 &= \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2, \qquad (5) \\
\omega &> 0, \quad \alpha_i \geq 0, \quad i = 1, \ldots, q,
\end{aligned}
$$

  which is referred to as ARCH(q).

- Conditions (6) make sure that σ_t^2 may not become negative.

[1] Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica* 50, 987-1007.

- σ_t^2 in (5) is the conditional variance of ϵ_t, given I_{t-1}.

- To find the unconditional variance, take expectations in (5),

$$\mathsf{E}(\sigma_t^2) = \mathsf{E}(\epsilon_t^2) = \omega + \sum_{i=1}^{q} \alpha_i \mathsf{E}(\epsilon_{t-i}^2),$$

so that

$$\mathsf{E}(\sigma_t^2) = \mathsf{E}(\epsilon_t^2) = \frac{\omega}{1 - \alpha_1 - \alpha_2 - \cdots - \alpha_q}.$$

- This makes sense only if

$$\sum_{i=1}^{q} \alpha_i < 1, \qquad (6)$$

which turns out to be the condition for the finiteness of the variance in the ARCH(q) model, and is often referred to as the stationarity condition.

- If the covariance stationarity condition (6) is not satisfied, this does *not* imply that the process is not (strictly) stationary.

- It means that the unconditional distribution has no finite second moment.

- It has been shown that the ARCH process (even with normal innovations) generates marginal (unconditional) distributions with tails decaying as a power law, i.e., for some $\gamma > 0$,

$$\Pr(|\epsilon_t| > x) \simeq cx^{-\gamma}, \quad \text{as } x \to \infty,$$

  so that moments of ϵ_t exist only of order smaller than γ.

- It may happen that the coefficients of the ARCH equation are so large that $\gamma < 2$.

- The (weaker) condition for strict stationarity will be briefly considered when discussing generalized ARCH (GARCH) models.

- Several further properties of the model can best be illustrated by means of the ARCH(1) specification, given by

$$\sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2. \tag{7}$$

- We first calculate the fourth moment of the process,

$$\mathsf{E}(\epsilon_t^4) = \mathsf{E}(\eta_t^4 \sigma_t^4) = \mathsf{E}(\eta_t^4)\mathsf{E}(\sigma_t^4) = 3\mathsf{E}(\sigma_t^4). \tag{8}$$

- Squaring (7),

$$\begin{aligned}
\sigma_t^4 &= (\omega + \alpha_1 \epsilon_{t-1}^2)^2 = \omega^2 + 2\omega\alpha_1\epsilon_{t-1}^2 + \alpha_1^2 \epsilon_{t-1}^4 \\
\mathsf{E}(\sigma_t^4) &= \omega^2 + 2\omega\alpha_1 \mathsf{E}(\epsilon_t^2) + \alpha_1^2 \mathsf{E}(\epsilon_t^4) \\
&= \omega^2 + \frac{2\omega^2\alpha_1}{1-\alpha_1} + 3\alpha_1^2 \mathsf{E}(\sigma_t^4) \\
\mathsf{E}(\sigma_t^4) &= \frac{1}{1-3\alpha_1^2}\left[\omega^2 + \frac{2\omega^2\alpha_1}{1-\alpha_1}\right] = \frac{\omega^2(1+\alpha_1)}{(1-\alpha_1)(1-3\alpha_1^2)},
\end{aligned}$$

which makes sense only if $3\alpha^2 < 1$, which is the condition for the finiteness of the fourth moment.

- In this case, from (8)

$$\mathsf{E}(\epsilon_t^4) = \frac{3\omega^2(1+\alpha_1)}{(1-\alpha_1)(1-3\alpha_1^2)}, \tag{9}$$

and the kurtosis of the unconditional distribution is, with $\mathsf{E}(\epsilon_t^2) = \omega/(1-\alpha_1)$,

$$\begin{aligned}
\frac{\mathsf{E}(\epsilon_t^4)}{\mathsf{E}^2(\epsilon_t^2)} &= \frac{3\omega^2(1+\alpha_1)(1-\alpha_1)^2}{\omega^2(1-\alpha_1)(1-3\alpha_1^2)} \\
&= \frac{3(1-\alpha_1)(1+\alpha_1)}{1-3\alpha_1^2} \\
&= \frac{3(1-\alpha_1^2)}{1-3\alpha_1^2} > 3.
\end{aligned}$$

- The ACF of the squared process,

$$\varrho(\tau) = \mathsf{Corr}(\epsilon_t^2, \epsilon_{t-\tau}^2) = \frac{\mathsf{E}(\epsilon_t^2 \epsilon_{t-\tau}^2) - \mathsf{E}^2(\epsilon_t^2)}{\mathsf{E}(\epsilon_t^4) - \mathsf{E}^2(\epsilon_t^2)}, \tag{10}$$

which is well–defined for $3\alpha_1^2 < 1$, is also of interest.

- We find

$$\begin{aligned}
\mathsf{E}(\epsilon_t^2 \epsilon_{t-\tau}^2) &= \mathsf{E}(\epsilon_{t-\tau}^2 \eta_t^2 \underbrace{(\omega + \alpha_1 \epsilon_{t-1}^2)}_{=\sigma_t^2}) \\
&= \omega \mathsf{E}(\epsilon_t^2) + \alpha_1 \mathsf{E}(\epsilon_{t-\tau}^2 \epsilon_{t-1}^2) \\
&= \mathsf{E}^2(\epsilon_t^2)(1 - \alpha_1) + \alpha_1 \mathsf{E}(\epsilon_{t-\tau}^2 \epsilon_{t-1}^2) \\
&= \mathsf{E}^2(\epsilon_t^2) + \alpha_1 [\mathsf{E}(\epsilon_{t-\tau}^2 \epsilon_{t-1}^2) - \mathsf{E}^2(\epsilon_t^2)] \\
\mathsf{E}(\epsilon_t^2 \epsilon_{t-\tau}^2) - \mathsf{E}^2(\epsilon_t^2) &= \alpha_1 [\mathsf{E}(\epsilon_{t-\tau}^2 \epsilon_{t-1}^2) - \mathsf{E}^2(\epsilon_t^2)],
\end{aligned}$$

which implies $\varrho(\tau) = \alpha_1 \varrho(\tau - 1)$.

- For $\tau = 1$, we have

$$\mathsf{E}(\epsilon_t^2 \epsilon_{t-1}^2) - \mathsf{E}^2(\epsilon_t^2) = \alpha_1[\mathsf{E}(\epsilon_t^4) - \mathsf{E}^2(\epsilon_t^2)],$$

so

$$\varrho(\tau) = \alpha^\tau. \tag{11}$$

GARCH Models

- In practice, pure ARCH(q) processes are rarely used, since for an adequate fit a large number of lags is usually required.

- A more parsimonious formalization is provided by the **G**eneralized ARCH (GARCH) process, as proposed by Bollerslev (1986) and Taylor (1986).[2]

- The GARCH(p, q) model generalizes (5) to

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2. \tag{12}$$

- To make sure that the variance is positive, Bollerslev (1986) imposed that

$$\omega > 0; \quad \alpha_i \geq 0, i = 1, \ldots, q; \quad \beta_i \geq 0, i = 1, \ldots, p. \tag{13}$$

[2] T. Bollerslev (1986): Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics* 31, 307-327. S. J. Taylor (1986): Modelling Financial Time Series, Wiley.

- These conditions are sufficient but can be weakened for models where one of the orders is larger than unity (see below). Conditions (13) are necessary and sufficient for guaranteeing a positive variance process in pure ARCH processes and the GARCH(1,1) process, however.

- Similar to the ARCH(q) process, we can calculate the unconditional variance of process as

$$\mathsf{E}(\sigma_t^2) = \mathsf{E}(\epsilon_t^2) = \frac{\omega}{1 - \sum_{i=1}^{q} \alpha_i - \sum_{i=1}^{p} \beta_i}, \tag{14}$$

provided the (covariance) stationarity condition

$$\sum_{i=1}^{q} \alpha_i + \sum_{i=1}^{p} \beta_i < 1 \tag{15}$$

is satisfied.

- To characterize the correlation structure of the squared process, define the prediction error

$$u_t = \epsilon_t^2 - \mathsf{E}(\epsilon_t^2 | I_{t-1}) = \epsilon_t^2 - \sigma_t^2. \tag{16}$$

- $u_t = \epsilon_t^2 - \sigma_t^2 = (\eta_t^2 - 1)\sigma_t^2$ is white noise but not strict white noise, since it is uncorrelated but not independent.

- Substituting (17) for σ_t^2 into (12) results in

$$\epsilon_t^2 = \omega + \sum_{i=1}^{\max\{p,q\}} (\alpha_i + \beta_i)\epsilon_{t-i}^2 - \sum_{i=1}^{p} \beta_i u_{t-i} + u_t, \qquad (17)$$

where $\alpha_i = 0$ for $i > q$ and $\beta_i = 0$ for $i > p$.

- Equation (17) is an ARMA($\max\{p,q\}, p$) representation for the *squared* process $\{\epsilon_t^2\}$, which characterizes its autocorrelations.

- The ARMA representation can also be used to explicitly calculate the autocorrelations.

- For example, the ARMA(1,1) representation of the GARCH(1,1) process is

$$\epsilon_t^2 = \omega + (\alpha_1 + \beta_1)\epsilon_{t-1}^2 + u_t - \beta_1 u_t. \qquad (18)$$

- Recall that the ACF of the ARMA(1,1) process

$$Y_t = \phi Y_{t-1} + \theta\epsilon_{t-1} + \epsilon_t$$

is

$$\text{Corr}(Y_t, Y_{t-\tau}) = \phi^{\tau-1}\frac{(\phi+\theta)(1+\phi\theta)}{1+2\theta\phi+\theta^2}.$$

- Plugging in $\alpha_1 + \beta_1$ for ϕ and $-\beta_1$ for θ gives the ACF of the squares of a GARCH(1,1) process as

$$\varrho(\tau) = (\alpha_1+\beta_1)^{\tau-1}\frac{\alpha_1(1-\alpha_1\beta_1-\beta_1^2)}{1-2\alpha_1\beta_1-\beta_1^2},$$

provided the fourth moment is finite (see below).

- The GARCH(1,1) process is most often applied in practice.

- To find the moments of this process, it is convenient to write

$$\begin{aligned}\sigma_t^2 &= \omega + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 = \omega + (\alpha_1 \eta_{t-1}^2 + \beta_1)\sigma_{t-1}^2 \\ &= \omega + c_{t-1}\sigma_{t-1}^2, \quad c_t = \alpha_1 \eta_t^2 + \beta_1.\end{aligned}$$

- Note that σ_{t-1}^2 is determined based on the information up to time $t-2$.

- c_{t-1} depends on η_{t-1}.

- Thus c_{t-1} and σ_{t-1}^2 are independent, and

$$\mathsf{E}(c_{t-1}^m \sigma_t^n) = \mathsf{E}(c_{t-1}^m)\mathsf{E}(\sigma_t^n) \tag{19}$$

  for all m and n.

- We have

$$\mathsf{E}(c_t) = \alpha_1 + \beta_1, \quad \mathsf{E}(c_t^2) = \mathsf{E}(\alpha_1^2 \eta_t^4 + 2\alpha_1\beta_1\eta_t^2 + \beta_1^2) = 3\alpha_1^2 + 2\alpha_1\beta_1 + \beta_1^2.$$

- Since $\mathsf{E}(\sigma_t^2) = \frac{\omega}{1-\alpha-\beta} = \omega/(1 - \mathsf{E}(c_t))$,

$$\begin{aligned}
\mathsf{E}(\sigma_t^4) &= \omega^2 + 2\omega \mathsf{E}(c_t)\mathsf{E}(\sigma_t^2) + \mathsf{E}(c_t^2)\mathsf{E}(\sigma_t^4) \\
\mathsf{E}(\sigma_t^4) &= \frac{\omega^2(1 + \mathsf{E}(c_t))}{(1 - \mathsf{E}(c_t))(1 - \mathsf{E}(c_t^2))} \\
&= \frac{\omega^2(1 + \alpha_1 + \beta_1)}{(1 - \alpha_1 - \beta_1)(1 - 3\alpha_1^2 - 2\alpha_1\beta_1 - \beta_1^2)},
\end{aligned}$$

where $\mathsf{E}(c_t^2) = 3\alpha_1^2 + 2\alpha_1\beta_1 + \beta_1^2 < 1$ is the condition for the existence of the fourth moment.

- The kurtosis is then

$$\begin{aligned}
\frac{\mathsf{E}(\epsilon_t^4)}{\mathsf{E}^2(\epsilon_t^2)} &= 3\frac{(1 - \alpha_1 - \beta_1)(1 + \alpha_1 + \beta_1)}{1 - 3\alpha_1^2 - 2\alpha_1\beta_1 - \beta_1^2} \\
&= 3 + \frac{6\alpha_1^2}{1 - 3\alpha_1^2 - 2\alpha_1\beta_1 - \beta_1^2}.
\end{aligned}$$

- To illustrate why GARCH(1,1) typically fits better than even a high–order ARCH(q), we write in lag–operator form and invert (assuming $\beta_1 < 1$)

$$\begin{aligned}
\sigma_t^2 &= \omega + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 && (20)\\
(1-\beta_1 L)\sigma_t^2 &= \omega + \alpha_1 \epsilon_{t-1}^2 && (21)\\
\sigma_t^2 &= \frac{\omega}{1-\beta_1} + \frac{\alpha_1 \epsilon_{t-1}^2}{1-\beta_1 L} && (22)\\
&= \frac{\omega}{1-\beta_1} + \alpha_1 \sum_{i=1}^{\infty} \beta_1^{i-1} \epsilon_{t-i}^2. && (23)
\end{aligned}$$

- This shows that GARCH(1,1) is ARCH(∞) with geometrically declining lag structure, i.e., $\sigma_t^2 = \tilde{\omega} + \sum_{i=1}^{\infty} \psi_i \epsilon_{t-i}^2$, with $\psi_i = \alpha_1 \beta_1^{i-1}$.

- The declining lag structure is reasonable as it implies that the impact of more recent shocks on the current variance is larger than that of earlier shocks.

- The ARCH(∞) representation (20) shows that α_1 can be interpreted as a *reaction parameter*, as it measures the reactiveness of the conditional variance to a shock in the previous period, i.e., the immediate impact of a unit shock on the next period's conditional variance.

- Parameter β_1, on the other hand, is a *memory parameter* which measures the memory in the variance process. E.g., if β_1 is small, β_1^i tends to zero very rapidly with i, and the *direct* impact of a shock on future conditional variances dies out soon.

Note on the nonnegativity conditions (13)

- We can use lag–operator notation to write the GARCH model as

$$\beta(L)\sigma_t^2 = \omega + \alpha(L)\epsilon_t^2,$$

where

$$\begin{aligned} \beta(L) &= 1 - \beta_1 L - \beta_2 L^2 - \cdots - \beta_p L^p \\ \alpha(L) &= \alpha_1 L + \alpha_2 L^2 + \cdots + \alpha_q L^q. \end{aligned}$$

Inverting gives the ARCH(∞)[3]

$$\sigma_t^2 = \frac{\omega}{1 - \sum_i \beta_i} + \frac{\alpha(L)}{\beta(L)}\epsilon_t^2 = \frac{\omega}{1 - \sum_i \beta_i} + \sum_{i=1}^{\infty} \psi_i \epsilon_{t-i}^2.$$

[3] This requires that $\beta(z)$ has all roots outside the unit circle.

- For σ_t^2 to remain positive with probability 1, we observe that it is necessary and sufficient that

$$\frac{\omega}{1-\sum_i \beta_i} > 0, \quad \psi_i \geq 0 \text{ for all } i.$$

- These restrictions are weaker than (13) except for the pure ARCH(q) and the GARCH(1,1).

- The simplest case is the GARCH(1,2),

$$\begin{aligned}
\sigma_t^2 &= \omega + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 \epsilon_{t-2}^2 + \beta_1 \sigma_{t-1}^2 \\
(1-\beta_1 L)\sigma_t^2 &= \omega + (\alpha_1 L + \alpha_2 L^2)\epsilon_t^2 \\
\sigma_t^2 &= \frac{\omega}{1-\beta_1} + \left(\frac{\alpha_1 L}{1-\beta_1 L} + \frac{\alpha_2 L^2}{1-\beta_1 L}\right)\epsilon_t^2 \\
&= \frac{\omega}{1-\beta_1} + \left(\alpha_1 \sum_{i=1}^{\infty} \beta_1^{i-1} L^i + \alpha_2 \sum_{i=2}^{\infty} \beta_1^{i-2} L^i\right)\epsilon_t^2 \\
&= \frac{\omega}{1-\beta_1} + \alpha_1 \epsilon_{t-1}^2 + \sum_{i=2}^{\infty} (\alpha_1 \beta_1^{i-1} + \alpha_2 \beta_1^{i-2})\epsilon_{t-i}^2 \\
&= \frac{\omega}{1-\beta_1} + \alpha_1 \epsilon_{t-1}^2 + \sum_{i=2}^{\infty} \beta_1^{i-2}(\alpha_1 \beta_1 + \alpha_2)\epsilon_{t-i}^2
\end{aligned}$$

- Thus

$$\begin{aligned}
\psi_1 &= \alpha_1 \\
\psi_k &= \beta_1^{k-2}(\alpha_1 \beta_1 + \alpha_2), \quad k \geq 2.
\end{aligned}$$

- This gives rise to the set of necessary and sufficient conditions

$$
\begin{aligned}
\omega &> 0 \\
\alpha_1 &\geq 0 \\
1 > \beta_1 &\geq 0 \\
\alpha_1\beta_1 + \alpha_2 &\geq 0.
\end{aligned}
$$

- α_2 may be negative if $\alpha_1 > 0$ and $\beta_1 > 0$.

- For the most frequently applied GARCH(1,1) process, however, the nonnegativity constraints $\omega > 0$, $\alpha, \beta \geq 0$ are necessary.

Testing for GARCH

- The tests have to be applied to the residuals $\{\widehat{\epsilon_t}\}_{t=1}^{T}$ of a model for the conditional mean, which may include exogenous factors time series components (such as ARMA), or just a constant.

- The Ljung–Box–Pierce statistic for the autocorrelations of the squares,

$$Q^\star = T(T+2)\sum_{\tau=1}^{K}\frac{\widehat{\rho}_{\widehat{\epsilon}^2}(\tau)^2}{T-\tau} \overset{asy}{\sim} \chi^2(K). \tag{24}$$

- Engle (1982) derived a Lagrange multiplier test which works as follows.

- Run the regression with q lags

$$\epsilon_t^2 = b_0 + b_1\widehat{\epsilon}_{t-1}^2 + \cdots + b_q\widehat{\epsilon}_{t-q}^2 + u_t. \tag{25}$$

- Under H_0 of no ARCH effects (conditional homoskedasticity), the test statistic

$$LM = TR^2 \overset{asy}{\sim} \chi^2(q), \tag{26}$$

where T is the sample size and R^2 is the coefficient of determination obtained from the regression (25).

- The test has to be applied to the residuals of a model for the conditional mean (which may include exogenous factors, time series components, or just a constant).

Estimation

- GARCH models are most frequently estimated by conditional maximum likelihood.

- To illustrate, suppose we want to estimate an AR(1)–GARCH(1,1) model for returns r_t.

- That is, the conditional mean of the time series is described by an AR(1), and the conditional variance is driven by GARCH(1,1).

- If we assume conditional normality, the model is

$$\begin{aligned} r_t &= c + \phi r_{t-1} + \epsilon_t, \quad |\phi| < 1 && (27) \\ \epsilon_t &= \eta_t \sigma_t, \quad \eta_t \overset{iid}{\sim} \mathsf{N}(0,1) && (28) \\ \sigma_t^2 &= \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2 && (29) \\ \omega &> 0, \quad \alpha, \beta \geq 0. && (30) \end{aligned}$$

- The parameter vector is $\theta = [\theta_1, \theta_2]$, where $\theta_1 = [c, \phi]$ is the conditional mean part, and $\theta_2 = [\omega, \alpha, \beta]$ is the GARCH–part.

- We observe a stretch of length T, $\{r_t\}_{t=1}^T$, and a presample value r_0 (i.e., the first observation of our original sample).

- From the ARMA part, for a given value of θ_1, $\widehat{\theta}_1 = (\widehat{c}, \widehat{\phi})$, we calculate

$$\widehat{\epsilon}_t = r_t - \widehat{c} - \widehat{\phi} r_{t-1}, \quad t = 1, \ldots, T. \tag{31}$$

- The conditional log–likelihood function, $\log L(\theta)$, is then given by

$$\log L(\widehat{\theta}) = \sum_{t=1}^{T} \ell_t(\widehat{\theta}), \tag{32}$$

where, under conditional normality,

$$\ell_t(\widehat{\theta}) = -\frac{1}{2} \log \widehat{\sigma}_t^2 - \frac{1}{2} \frac{\widehat{\epsilon}_t^2}{\widehat{\sigma}_t^2}, \quad t = 1, \ldots, T, \tag{33}$$

and, for given $\widehat{\theta}_2 = (\widehat{\omega}, \widehat{\alpha}, \widehat{\beta})$,

$$\widehat{\sigma}_t^2 = \widehat{\omega} + \widehat{\alpha}\widehat{\epsilon}_{t-1}^2 + \widehat{\beta}\widehat{\sigma}_{t-1}^2, \quad t = 1, \ldots, T. \tag{34}$$

- To start the GARCH recursion (34), we need initial values $\widehat{\sigma}_0^2$ and $\widehat{\epsilon}_0^2$.

- One possibility is to set these equal to their unconditional values estimated from the sample at hand, i.e.,

$$\widehat{\sigma}_0^2 = \widehat{\epsilon}_0^2 = \frac{1}{T}\sum_{t=1}^{T}\widehat{\epsilon}_t^2, \tag{35}$$

with $\widehat{\epsilon}_t$, $t = 1, \ldots, T$, given by (31).

- Alternatively, we could treat $\widehat{\sigma}_0^2$ as an additional parameter to estimate, and estimate $\widehat{\epsilon}_0$ via the difference between r_0 and its unconditional mean implied by the AR(1), i.e., $\mathsf{E}(r_0) = c/(1-\phi)$.

- In practice, GARCH models are typically applied to sufficiently long time series, so that the choice of the initialization has negligible impact on the results.

- We then maximize (32) with respect to θ to obtain the maximum likelihood estimator (MLE) $\widehat{\theta}_{ML}$.

- Following standard large sample theory for the MLE, inference (e.g., calculation standard errors) is based on

$$\widehat{\theta}_{ML} \overset{approx}{\sim} \text{Normal}(\theta, I(\widehat{\theta}_{ML})^{-1}), \tag{36}$$

where

$$I(\widehat{\theta}_{ML}) = -\frac{\partial \log L(\widehat{\theta}_{ML})}{\partial\theta\partial\theta'} = -\sum_{t=1}^{T} \frac{\partial \ell_t(\widehat{\theta}_{ML})}{\partial\theta\partial\theta'} \tag{37}$$

is the negative of the Hessian matrix of the log–likelihood function, evaluated at the MLE.

- The derivatives in (37) can be calculated analytically or numerically.

- The Gaussian assumption for η_t often appears to be unreasonable.

- A frequently employed alternative is Student's t, in which case the density of η_t is

$$f(\eta_t;\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma(\nu/2)\sqrt{(\nu-2)\pi}}\left(1+\frac{\eta_t^2}{\nu-2}\right)^{-(\nu+1)/2},$$

where $\nu > 2$ is the degrees of freedom parameter and controls the thickness of the tails.

- Note that ν is a free parameter of the model that is estimated simultaneously with the other parameters from the data.

Fitting GARCH Models

- To illustrate typical results, we fit model

$$
\begin{aligned}
r_t &= \mu + \epsilon_t \\
\epsilon_t &= \eta_t \sigma_t, \quad \eta_t \overset{iid}{\sim} \mathsf{N}(0,1) \\
\sigma_t^2 &= \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2
\end{aligned}
$$

  to various stock index series.

- Parameter estimates are reported in Table 1.

Table 1: GARCH(1,1) estimates for various stock return series, approx. 1990–2010

Series	$\widehat{\omega}$	$\widehat{\alpha}_1$	$\widehat{\beta}_1$	$\widehat{\alpha}_1 + \widehat{\beta}_1$
S&P 500	0.0077 (0.0017)	0.0655 (0.0067)	0.9284 (0.0072)	0.9939
DAX	0.0355 (0.0053)	0.0918 (0.0089)	0.8910 (0.0099)	0.9828
FTSE	0.0113 (0.0025)	0.0856 (0.0081)	0.9059 (0.0087)	0.9915
CAC 40	0.0290 (0.0054)	0.0851 (0.0085)	0.9001 (0.0097)	0.9852

- Simple diagnostics can be based on the sequence of standardized residuals,

$$\widehat{\eta}_t = \frac{\widehat{\epsilon}_t}{\widehat{\sigma}_t}, \quad t = 1, \ldots, T. \tag{38}$$

- This sequence should behave like an iid sequence from the presumed innovation distribution.

- In particular, the GARCH model should capture all the conditional heteroskedasticity.

- Thus, sequence (38) should not display any conditional heteroskedasticity.

- This can be checked visually by plotting the SACF of the absolute or squared residuals, or by calculating test statistics for conditional heteroskedasticity, as discussed above.

- If the innovations have been assumed normal, we can apply normality tests to (38).

Table 2: Kurtosis of raw returns and residuals (38)

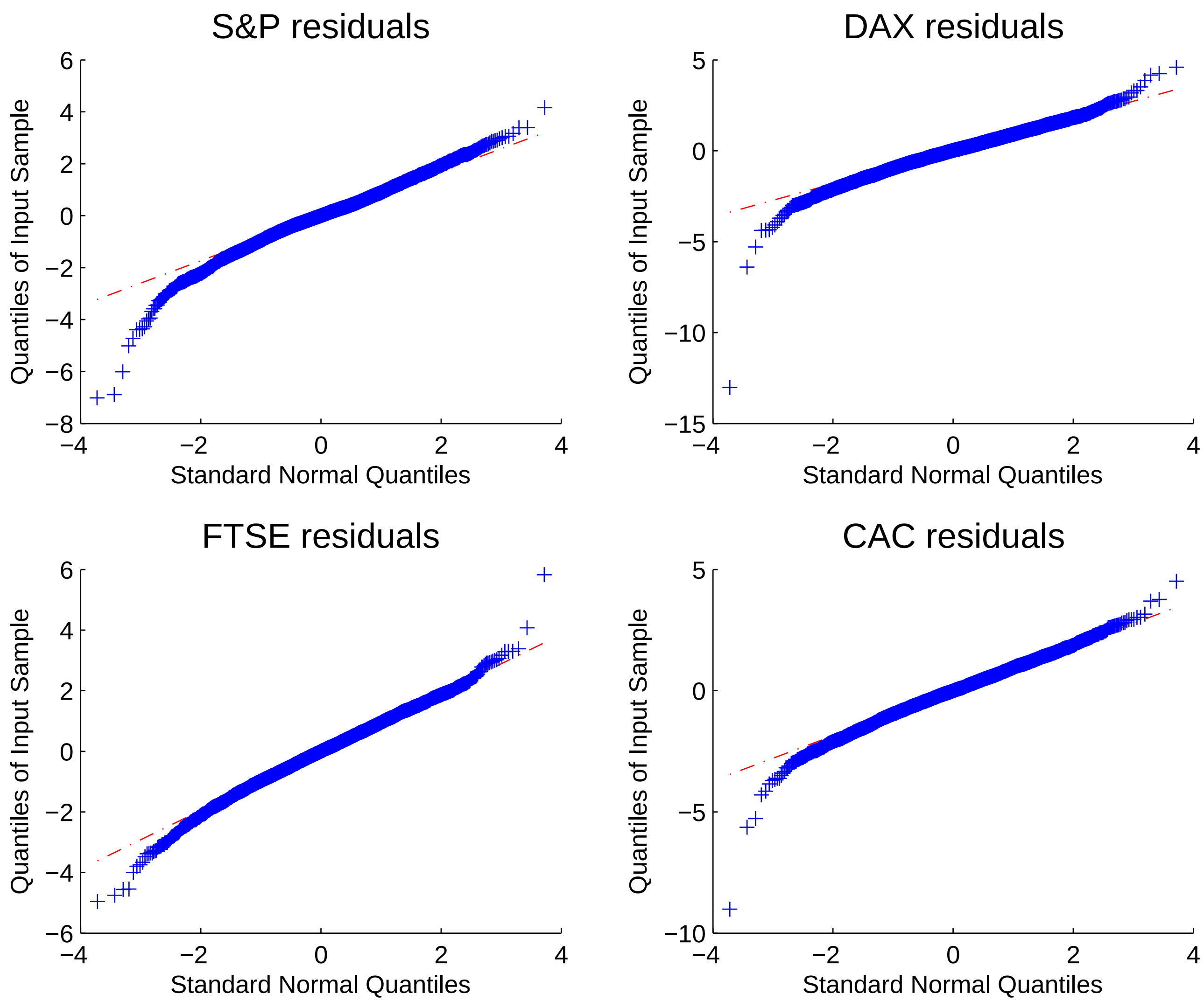
	S&P 500	DAX	FTSE	CAC 40
raw returns	12.1307	8.0553	9.6318	7.8069
residuals (38)	4.8993	9.6475	3.8232	4.9332

- We observe that GARCH captures part of the excess kurtosis in the unconditional distribution.

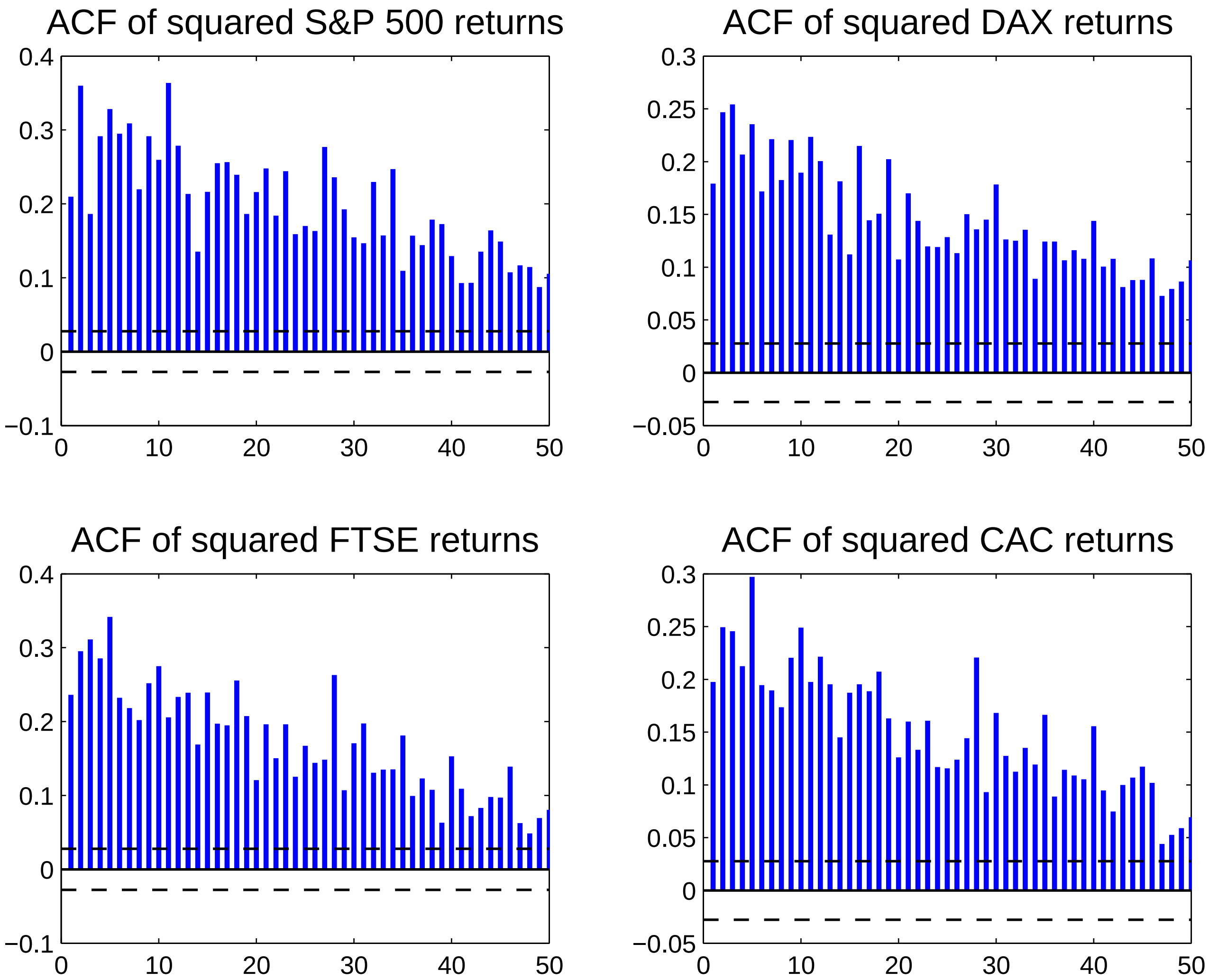
- (The number for the DAX is due to the Gorbatschow-Putsch in August 1991.)

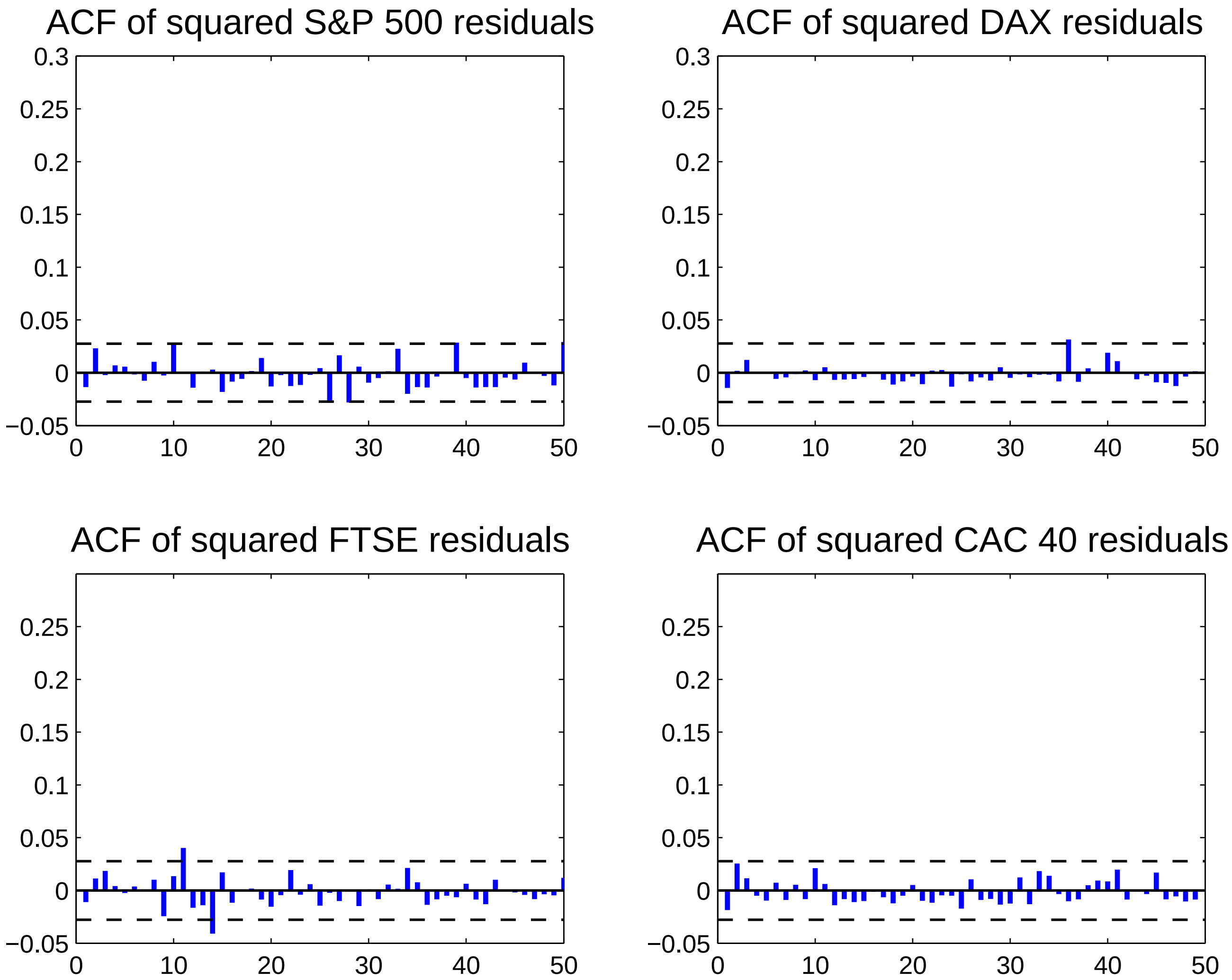
- However, the kurtosis of the standardized residuals (38) is still significantly different from the Gaussian value.

- That a leptokurtic (fat–tailed) innovation distribution may be appropriate.

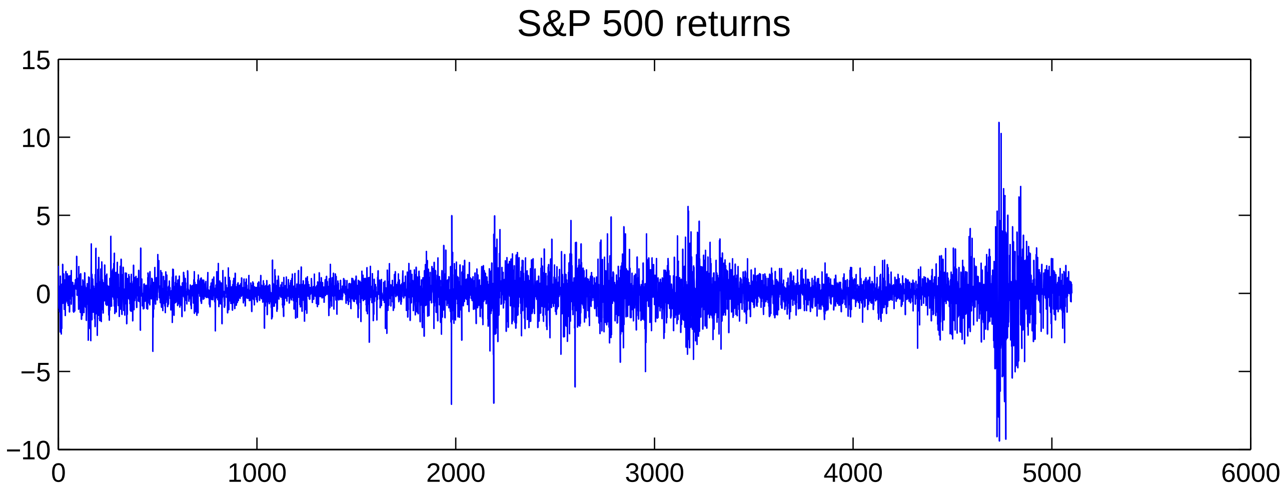


The Q–Q plots of the $\{\widehat{\eta}_t\}$ also indicate a fatter tailed innovation density.

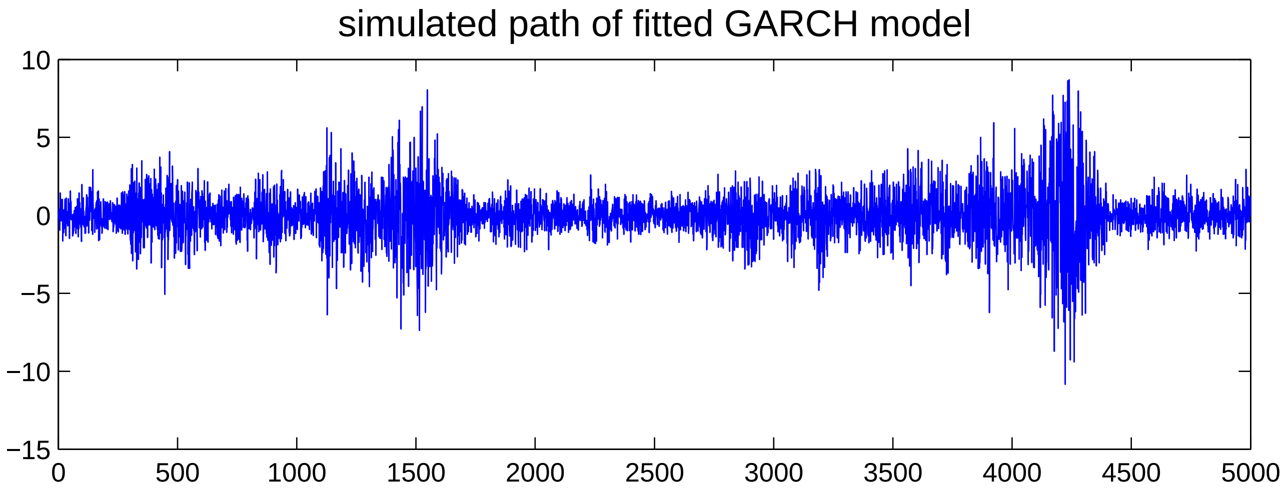

ACF of squared S&P 500 returns
ACF of squared DAX returns
ACF of squared FTSE returns
ACF of squared CAC returns
0.4
0.3
0.2
0.1
0
−0.1
0.25
0.15
0.05
−0.05
0
10
20
30
40
50



This applies to the standardized residuals (38). GARCH(1,1) appears to be sufficient.



S&P 500 returns
15
10
5
0
−5
−10
0
1000
2000
3000
4000
5000
6000



simulated path of fitted GARCH model
10
5
0
−5
−10
−15
0
500
1000
1500
2000
2500
3000
3500
4000
4500
5000

Alternative Innovation Distributions

- In view of these results, it appears reasonable to replace the normal distribution of η_t in the GARCH(1,1) with a more flexible alternative that allows for *conditional* leptokurtosis.

- Two of the most popular candidates in this regard are the
  - Student's t
  - Generalized Error Distribution (GED)

- The unit–variance versions of these are given by

$$f(\eta_t;\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma(\nu/2)\sqrt{(\nu-2)\pi}}\left(1+\frac{\eta_t^2}{\nu-2}\right)^{-(\nu+1)/2}, \qquad (39)$$

and

$$f(\eta_t;p) = \frac{\lambda p}{2^{1/p+1}\Gamma(1/p)}\exp\left\{-\frac{|\lambda\eta_t|^p}{2}\right\}, \qquad (40)$$

where $\lambda = 2^{1/p}\sqrt{\Gamma(3/p)/\Gamma(1/p)}$.

Covariance Stationarity and Unconditional Variance for General Innovation Distributions

- In the GARCH(p, q),

$$\begin{aligned} \epsilon_t &= \eta_t \sigma_t \qquad (41) \\ \sigma_t^2 &= \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2, \end{aligned}$$

to find the unconditional variance, we take expectations on both sides,

$$\mathsf{E}(\sigma_t^2) = \omega + \sum_{i=1}^{q} \alpha_i \mathsf{E}(\epsilon_{t-i}^2) + \sum_{i=1}^{p} \beta_i \mathsf{E}(\sigma_{t-i}^2).$$

- If the innovations η_t in (41) have unit variance, $\mathsf{E}(\eta_t^2) = 1$, it follows that $\mathsf{E}(\epsilon_t^2) = \mathsf{E}(\eta_t^2 \sigma_t^2) = \mathsf{E}(\eta_t^2)\mathsf{E}(\sigma_t^2) = \mathsf{E}(\sigma_t^2)$, and so

$$\mathsf{E}(\epsilon_t^2) = \mathsf{E}(\sigma_t^2) = \frac{\omega}{1 - \sum_i \alpha_i - \sum_i \beta_i}, \qquad (42)$$

provided the second–order stationarity condition

$$\sum_i \alpha_i + \sum_i \beta_i < 1 \tag{43}$$

is satisfied.

- However, non–normal densities are not always applied in standardized (unit–variance) form.

- For example, the "conventional" Student's t is also often used and has density
$$f(\eta_t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma(\nu/2)\sqrt{\nu\pi}} \left(1 + \frac{\eta_t^2}{\nu}\right)^{-(\nu+1)/2},$$
which has (for $\nu > 2$)

$$\kappa_2 := \mathsf{E}(\eta_t^2) = \frac{\nu}{\nu - 2}.$$

- If, in general, $\mathsf{E}(\eta_t^2) = \kappa_2$, then (43) and (42) become

$$\kappa_2 \sum_i \alpha_i + \sum_i \beta_i < 1,$$

and

$$\mathsf{E}(\epsilon_t^2) = \kappa_2 \mathsf{E}(\sigma_t^2) = \frac{\kappa_2 \omega}{1 - \kappa_2 \sum_i \alpha_i - \sum_i \beta_i},$$

respectively.

Table 3: GARCH(1,1) estimates for various stock return series, January 1990 to October 2009

Student's t

Series	$\widehat{\omega}$	$\widehat{\alpha}_1$	$\widehat{\beta}_1$	$\widehat{\nu}$	$\widehat{\alpha}_1 + \widehat{\beta}_1$
CAC 40	0.0197 (0.0048)	0.0751 (0.0081)	0.9150 (0.0089)	11.2269 (1.4925)	0.9902
DAX	0.0154 (0.0039)	0.0852 (0.0092)	0.9096 (0.0093)	8.6085 (0.8904)	0.9948
FTSE	0.0103 (0.0025)	0.0797 (0.0084)	0.9122 (0.0090)	13.2518 (2.0502)	0.9918

GED

Series	$\widehat{\omega}$	$\widehat{\alpha}_1$	$\widehat{\beta}_1$	$\widehat{p}$	$\widehat{\alpha}_1 + \widehat{\beta}_1$
CAC 40	0.0240 (0.0054)	0.0787 (0.0087)	0.9090 (0.0098)	1.5772 (0.0430)	0.9878
DAX	0.0226 (0.0048)	0.0882 (0.0099)	0.9027 (0.0103)	1.4412 (0.0364)	0.9909
FTSE	0.0110 (0.0026)	0.0825 (0.0086)	0.9089 (0.0092)	1.6790 (0.0472)	0.9914

Table 4: Maximized log–likelihood values

	CAC 40	DAX	FTSE
Normal	–8088.5	–8180.9	–6798.8
Student's t	–8032.5	–8048.2	–6768.2
GED	–8048.6	–8085.1	–6779.0

Differences in log–likelihood

Student's t–Normal	56.0047	132.6939	30.6056
GED–Normal	39.8959	95.7972	19.8580

IGARCH and EWMA

- The finding that often $\alpha_1 + \beta_1 \approx 1$ has led to the suggestion of imposing the restriction
$$\alpha_1 + \beta_1 = 1,$$
which is referred to as IGARCH(1,1) (integrated GARCH), since there is a "unit root" in the GARCH polynomial.

- However, the analogy to integrated (unit root) processes is rather weak.

- In particular, IGARCH(1,1) processes are (strictly) stationary, although their second moment does not exist.

- Nelson (1990) has shown that the GARCH(1,1) is strictly stationary if
$$\mathsf{E}[\log(\alpha_1 \eta_t^2 + \beta_1)] < 0.$$

- By Jensen's inequality, for the IGARCH(1,1),
$$\mathsf{E}[\log(\alpha_1 \eta_t^2 + \beta_1)] < \log \mathsf{E}(\alpha_1 \eta_t^2 + \beta_1) = \log 1 = 0.$$

- $\alpha_1 + \beta_1$ may be even larger than unity. For example, the ARCH(1) process with $\alpha_1 = 3$ is stationary, although extremely fat–tailed.

- A special case of an IGARCH model (with zero intercept) is the exponentially weighted moving average (EWMA) popularized by RiskMetrics of J.P. Morgan, which is

$$\sigma_t^2 = (1-\lambda)\sum_{i=0}^{\infty} \lambda^i \epsilon_{t-1-i}^2 = (1-\lambda)\epsilon_{t-1}^2 + \lambda\sigma_{t-1}^2, \quad 0 < \lambda < 1, \quad (44)$$

  with λ fixed at 0.94 for daily data.

- IGARCH and EWMA tend to be inferior in empirical applications, however.

Backtesting Predictive Densities of Nonlinear Time Series Models

- In the Gaussian GARCH model, series

$$\widehat{\eta}_t = \frac{\widehat{\epsilon}_t}{\widehat{\sigma}_t}, \quad t = 1, \ldots, T, \tag{45}$$

  should mimic an iid standard normally distributed series.

- Similarly, in a t or GED GARCH model, (45) should behave like an iid standard t or GED sequence.

- A frequently used technique to generate iid standard normal residuals is as follows.

- Calculate the series

$$u_t = F(r_t|I_{t-1}), \quad t = 1, \ldots, T, \tag{46}$$

where $F(\cdot|I_{t-1})$ is the *conditional* cumulative distribution function (cdf) of the return r_t implied by the model under consideration, based on information up to time $t-1$, I_{t-1}.

- For example, in a GARCH model with normal innovations, $F(\cdot|I_{t-1})$ is the normal cdf,

$$F(r|I_{t-1}) = \Phi\left(\frac{r_t - \widehat{\mu}_t}{\widehat{\sigma}_t}\right) = \frac{1}{\sqrt{2\pi}\widehat{\sigma}_t}\int_{-\infty}^{r} \exp\left\{-\frac{(\xi - \widehat{\mu}_t)^2}{2\widehat{\sigma}_t^2}\right\} d\xi, \tag{47}$$

where Φ is the standard normal cdf and $\widehat{\mu}_t$ and $\widehat{\sigma}_t^2$ are the conditional mean and variance implied by the estimated model, respectively.

- Computer programs do these calculations for most of the commonly used distributions.

- If the model is correctly specified, (46) is a series of iid uniform(0,1) variables. This is also known as the *Rosenblatt* transform.

- Subsequently, apply a second transformation, namely,

$$\{z_t\} = \Phi^{-1}(\{u_t\}), \tag{48}$$

where Φ^{-1} is the inverse of the standard normal cdf.

- For example,

$$\Phi^{-1}(.025) = -1.96, \quad \Phi^{-1}(.05) = -1.6449, \quad \Phi^{-1}(.5) = 0.$$

- If the model is correctly specified, (48) is a sequence of iid standard normal variables.

- This allows the use of standard and simple normality tests for correct specification of (conditional) skewness and kurtosis.

- Let $\widehat{s}$ and $\widehat{\kappa}$ be the sample skewness and kurtosis, respectively, i.e.,

$$\widehat{s} = \frac{T^{-1}\sum_t (z_t - \bar{z})^3}{\left\{T^{-1}\sum_t (z_t - \bar{z})^2\right\}^{3/2}}, \quad \widehat{\kappa} = \frac{T^{-1}\sum_t (z_t - \bar{z})^4}{\left\{T^{-1}\sum_t (z_t - \bar{z})^2\right\}^2}.$$

- Under normality,

$$\widehat{s} \overset{asy}{\sim} \text{Normal}(0, 6/T), \quad \widehat{\kappa} \overset{asy}{\sim} \text{Normal}(3, 24/T), \tag{49}$$

so

$$T\widehat{s}^2/6 \overset{asy}{\sim} \chi^2(1), \quad T(\widehat{\kappa} - 3)^2/24 \overset{asy}{\sim} \chi^2(1), \tag{50}$$

and the Jarque–Bera test

$$JB = T\widehat{s}^2/6 + T(\widehat{\kappa} - 3)^2/24 \overset{asy}{\sim} \chi^2(2). \tag{51}$$

- We can also test for absence of autocorrelation, zero mean and unit variance by means of likelihood ratio tests based on the Gaussian likelihood.

Economic Evaluation: Value–at–Risk (VaR)

- Both in industry and in academia, Value–at–Risk (VaR) is a widely employed measure to characterize the downside risk of a financial position.

- The VaR(ξ)

  – with *shortfall probability* ξ (typically a small number, e.g., $\xi = 0.01$ or 0.05)

  for a given horizon (typical a day or a week) is defined such that

  – over the next period (e.g., day or week), the probability that the portfolio suffers a loss larger than the VaR(ξ) is $100 \times \xi$%.

- Equivalently, with probability $1 - \xi$, our loss will not exceed the VaR(ξ).

- To be more precise, consider a time series of portfolio returns, r_t, and an associated series of ex–ante VaR measures with shortfall probability ξ, VaR$_t(\xi)$.

- The $\text{VaR}_t(\xi)$ implied by a model $\mathcal{M}$ is defined by

$$F_{t-1}^{\mathcal{M}}(\text{VaR}_t(\xi)) = \xi, \tag{52}$$

where $F_{t-1}^{\mathcal{M}}$ is the (conditional) cumulative distribution function (cdf) derived from model $\mathcal{M}$ using the information up to time $t-1$.

- Statistically, it is the ξ–quantile of the conditional return distribution.

- Under conditional normality, we have

$$\text{VaR}_t(\xi) = \mu_t + z_\xi \sigma_t,$$

where μ_t is the conditional mean of the return, z_ξ is the ξ–quantile of the standard normal distribution (e.g., $z_{0.01} = -2.3263$), and σ_t is the conditional standard deviation.

- A *violation* or *hit* is said to occur at time t if

$$r_t < \text{VaR}_t(\xi).$$

- For a nominal VaR shortfall probability ξ and a correctly specified VaR model, we expect $100 \times \xi$% of the observed return values to be violations (shortfalls).

- To test the models' suitability for calculating accurate ex–ante VaR measures, define the binary sequence

$$I_t = \begin{cases} 1, & \text{if} \quad r_t < \text{VaR}_t, \\ 0, & \text{if} \quad r_t \geq \text{VaR}_t. \end{cases} \tag{53}$$

- Then the empirical relative shortfall frequency is

$$\widehat{\xi} = x/T, \quad \text{where} \quad x = \sum_{t=1}^{T} I_t \tag{54}$$

  is the number of observed violations, and T is the number of forecasts evaluated.

- If $\widehat{\xi}$ is significantly higher (less) than ξ, then the model under study tends to underestimate (overestimate) the risk of the financial position.

- If the model is correctly specified, the hit sequence is a sample of size T from the Bernoulli distribution with parameter ξ, with pdf

$$p(I_t;\xi) = \xi^{I_t}(1-\xi)^{1-I_t}, \tag{55}$$

and the likelihood of the sample is

$$L(\xi) = \xi^{\sum_{t=1}^{T} I_t}(1-\xi)^{T-\sum_{t=1}^{T} I_t} = \xi^{x}(1-\xi)^{T-x}, \tag{56}$$

with log–likelihood

$$\log L(\xi) = x\log\xi + (T-x)\log(1-\xi). \tag{57}$$

- The maximum likelihood estimator is obtained via

$$\frac{\partial \log L(\xi)}{\partial\xi} = \frac{x}{\xi} - \frac{T-x}{1-\xi} = 0 \Rightarrow \widehat{\xi} = \frac{x}{T}. \tag{58}$$

- The likelihood ratio test statistic is two times the unrestricted log–likelihood,

$$\log L(\widehat{\xi}) = x \log(x/T) + (T - x) \log\{(T - x)/x\}, \tag{59}$$

minus the log–likelihood under the null that the actual shortfall probability is equal to the nominal shortfall probability ξ,

$$\log L(\widehat{\xi}) = x \log \xi + (T - x) \log(1 - \xi). \tag{60}$$

- The likelihood ratio test (LRT) statistic is

$$\text{LRT} = -2\{x \log(\xi/\widehat{\xi}) + (T - x) \log[(1 - \xi)/(1 - \widehat{\xi})]\} \overset{asy}{\sim} \chi^2(1). \tag{61}$$

One–step–ahead predictive densities

- First estimate the models over the (approximately) first ten years of data, i.e., the first 2500 observations.

- Then update the parameters (approximately) every month (i.e., 20 trading days) employing a moving window of data, i.e., using the most recent 2500 observations in the sample.

- We get, for each model and series, 2480 one–step–ahead predictive densities for the period January 2000 to October 2009.

Table 5: GARCH(1,1) density forecasts based on (48)

Gaussian GARCH(1,1)

Series	mean	var.	skewness	kurtosis	JB
CAC 40	-0.0562^{***}	1.0229	-0.304^{***}	4.014^{***}	144.5^{***}
DAX	-0.0567^{***}	1.0249	-0.317^{***}	3.945^{***}	133.7^{***}
FTSE	-0.0517^{**}	1.0221	-0.354^{***}	3.746^{***}	109.3^{***}

GED GARCH(1,1)

Series	mean	var.	skewness	kurtosis	JB
CAC 40	-0.0569^{***}	1.0162	-0.221^{***}	3.327^{***}	31.20^{***}
DAX	-0.0643^{***}	1.0152	-0.224^{***}	3.184^{*}	24.17^{***}
FTSE	-0.0538^{***}	1.0164	-0.275^{***}	3.238^{**}	37.06^{***}

Student's t GARCH(1,1)

Series	mean	var.	skewness	kurtosis	JB
CAC 40	-0.0584^{***}	1.0121	-0.185^{***}	3.097	15.11^{***}
DAX	-0.0636^{***}	1.0138	-0.187^{***}	2.983	14.55^{***}
FTSE	-0.0540^{***}	1.0144	-0.240^{***}	3.070	24.24^{***}

Asterisks *, **, and *** indicate significance at the 10%, 5% and 1% levels, respectively.

Table 6: GARCH(1,1) Value–at–Risk measures, reported is $100 \times \widehat{\xi}$

Gaussian GARCH(1,1)

Series	$\xi = 0.001$	0.0025	0.005	0.01	0.025	0.05	0.1
CAC 40	0.36^{***}	0.52^{**}	0.89^{**}	1.69^{***}	3.83^{***}	6.33^{***}	11.01^{*}
DAX	0.28^{**}	0.65^{***}	1.01^{***}	1.45^{**}	3.79^{***}	6.98^{***}	11.73^{***}
FTSE	0.60^{***}	0.77^{***}	1.25^{***}	2.02^{***}	3.95^{***}	6.37^{***}	10.56

GED GARCH(1,1)

Series	0.001	0.0025	0.005	0.01	0.025	0.05	0.1
CAC 40	0.32^{***}	0.36	0.60	1.17	3.67^{***}	6.25^{***}	11.33^{**}
DAX	0.20	0.28	0.65	1.13	3.31^{**}	6.98^{***}	12.50^{***}
FTSE	0.28^{**}	0.65^{***}	0.97^{***}	1.57^{***}	3.67^{***}	6.37^{***}	11.01^{*}

Student's t GARCH(1,1)

Series	0.001	0.0025	0.005	0.01	0.025	0.05	0.1
CAC 40	0.20	0.36	0.56	1.17	3.75^{***}	6.49^{***}	11.45^{**}
DAX	0.12	0.28	0.65	1.13	3.43^{***}	7.10^{***}	12.54^{***}
FTSE	0.24^{*}	0.65^{***}	0.85^{**}	1.61^{***}	3.83^{***}	6.57^{***}	11.41^{**}

Asterisks *, **, and *** indicate significance at the 10%, 5% and 1% levels, respectively, based on the test (61).

Conditional Skewness

- The results suggest that the innovations, in addition to being leptokurtic, are also skewed, which needs to be taken into account to deliver reliable density forecasts.

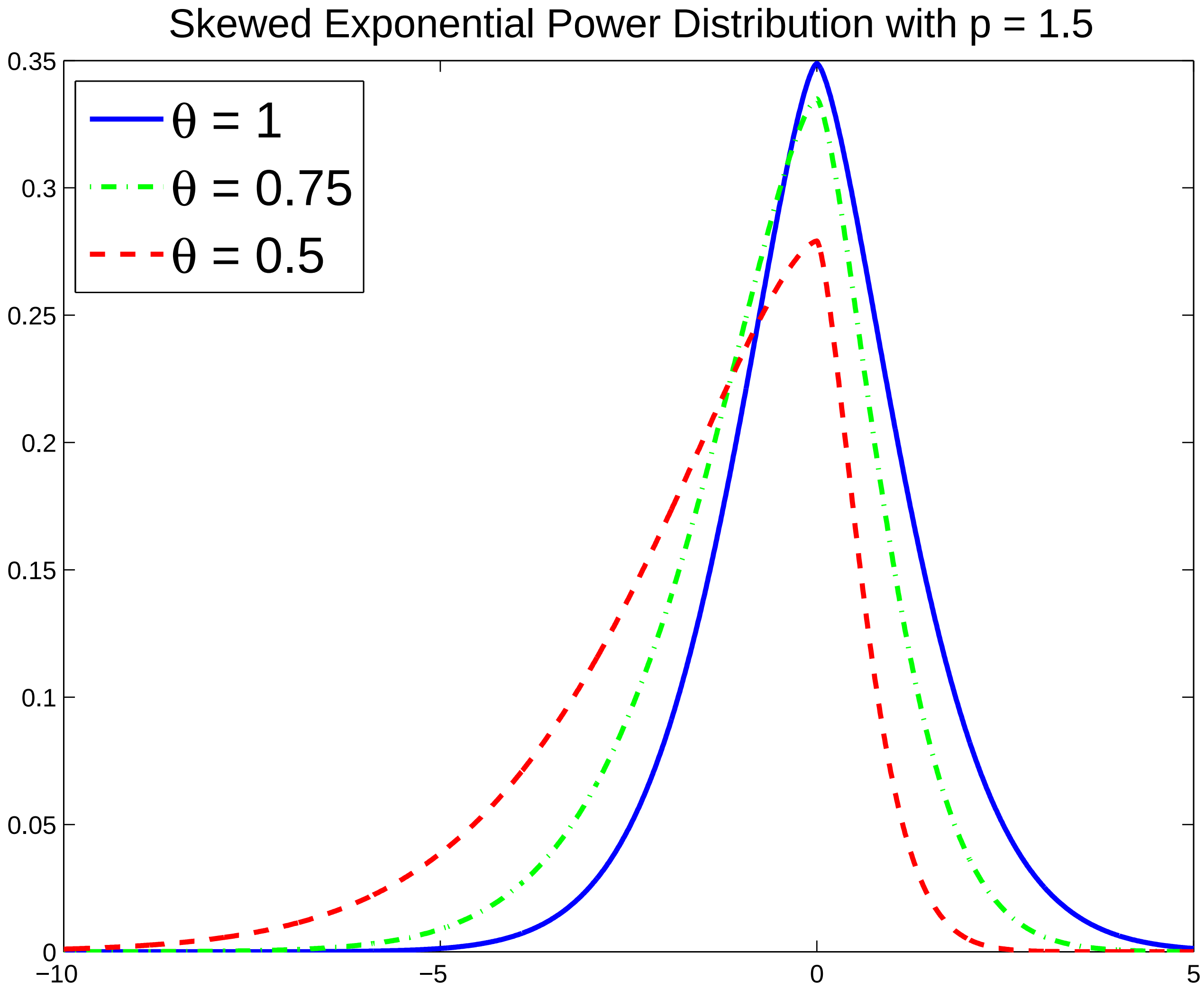
- Asymmetric versions of the GED and the t distributions have been proposed.

- Regarding the GED, the skewed exponential power (SEP) distribution of Fernandez, Osiewalski, and Steel (1995) has density

$$f(z;p,\theta) = \frac{\theta}{1+\theta^2}\frac{p}{2^{1/p}\Gamma(1/p)}\begin{cases}\exp\left\{-\frac{1}{2}\left(|z|\theta\right)^p\right\} & \text{if} \quad z<0 \\ \exp\left\{-\frac{1}{2}\left(\frac{z}{\theta}\right)^p\right\} & \text{if} \quad z\geq 0,\end{cases} \tag{62}$$

  where $\theta, p > 0$.

- This distribution nests the normal for $\theta = 1$ and $p = 2$. For $\theta < 1 (\theta > 1)$, the density is skewed to the left (right), and is fat–tailed for $p < 2$.


Skewed Exponential Power Distribution with p = 1.5
θ = 1
θ = 0.75
θ = 0.5
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
−10
−5
0
5

- Various skewed versions of the Student's t exist.

- A t version of (62) is the skewed t distribution proposed by Mittnik and Paolella (2000), which has density

$$f(z;\nu,p,\theta) = \frac{\theta}{1+\theta^2}\frac{p}{\nu^{1/p}B(\nu,1/p)}\begin{cases}\left(1+\frac{(|z|\theta)^p}{\nu}\right)^{-(\nu+1/p)} & \text{if} \quad z<0 \\ \left(1+\frac{(z/\theta)^p}{\nu}\right)^{-(\nu+1/p)} & \text{if} \quad z\geq 0,\end{cases} \tag{63}$$

where $\nu, p, \theta > 0$, and $B(\cdot,\cdot)$ is the beta function.

- In view of our earlier results that the (symmetric) t was somewhat better than the (symmetric) GED, we concentrate on the skewed t distribution (63).

Table 7: GARCH(1,1) estimates for various stock return series, January 1990 to October 2009

	Skewed Student's t					
Series	$\widehat{\omega}$	$\widehat{\alpha}_1$	$\widehat{\beta}_1$	$\widehat{\nu}$	$\widehat{\theta}$	$\widehat{p}$
CAC 40	0.0302	0.1307	0.9140	4.2942	0.9025	2.1988
	(0.0078)	(0.0142)	(0.0087)	(1.1377)	(0.0151)	(0.1483)
DAX	0.0237	0.1394	0.9076	3.2897	0.9005	2.2424
	(0.0064)	(0.0153)	(0.0092)	(0.6919)	(0.0143)	(0.1494)
FTSE	0.0167	0.1431	0.9119	3.8977	0.9100	2.3275
	(0.0043)	(0.0150)	(0.0085)	(1.0776)	(0.0148)	(0.1723)

- All the $\widehat{\theta}$s significantly different from 1.

Table 8: Maximized log–likelihood values

	CAC 40	DAX	FTSE
Normal	–8088.5	–8180.9	–6798.8
Student's t	–8032.5	–8048.2	–6768.2
GED	–8048.6	–8085.1	–6779.0
skewed t	–8013.1	–8024.8	–6749.3

Differences in log–likelihood

	CAC 40	DAX	FTSE
Student's t – Normal	56.0047	132.6939	30.6056
GED – Normal	39.8959	95.7972	19.8580
skew t – t	19.4212	23.3951	18.9364

- The 1% critical value of a $\chi^2(2)$ distribution is 9.2103.

Table 9: GARCH(1,1) density forecasts based on (48)

	skewed t GARCH(1,1)				
Series	mean	var.	skewness	kurtosis	JB
CAC 40	-0.0345^{*}	1.0042	-0.036	3.126	2.163
DAX	-0.0241	1.0131	-0.029	3.035	0.464
FTSE	-0.0189	1.0126	-0.088^{*}	3.067	3.630

Asterisks *, **, and *** indicate significance at the 10%, 5% and 1% levels, respectively.

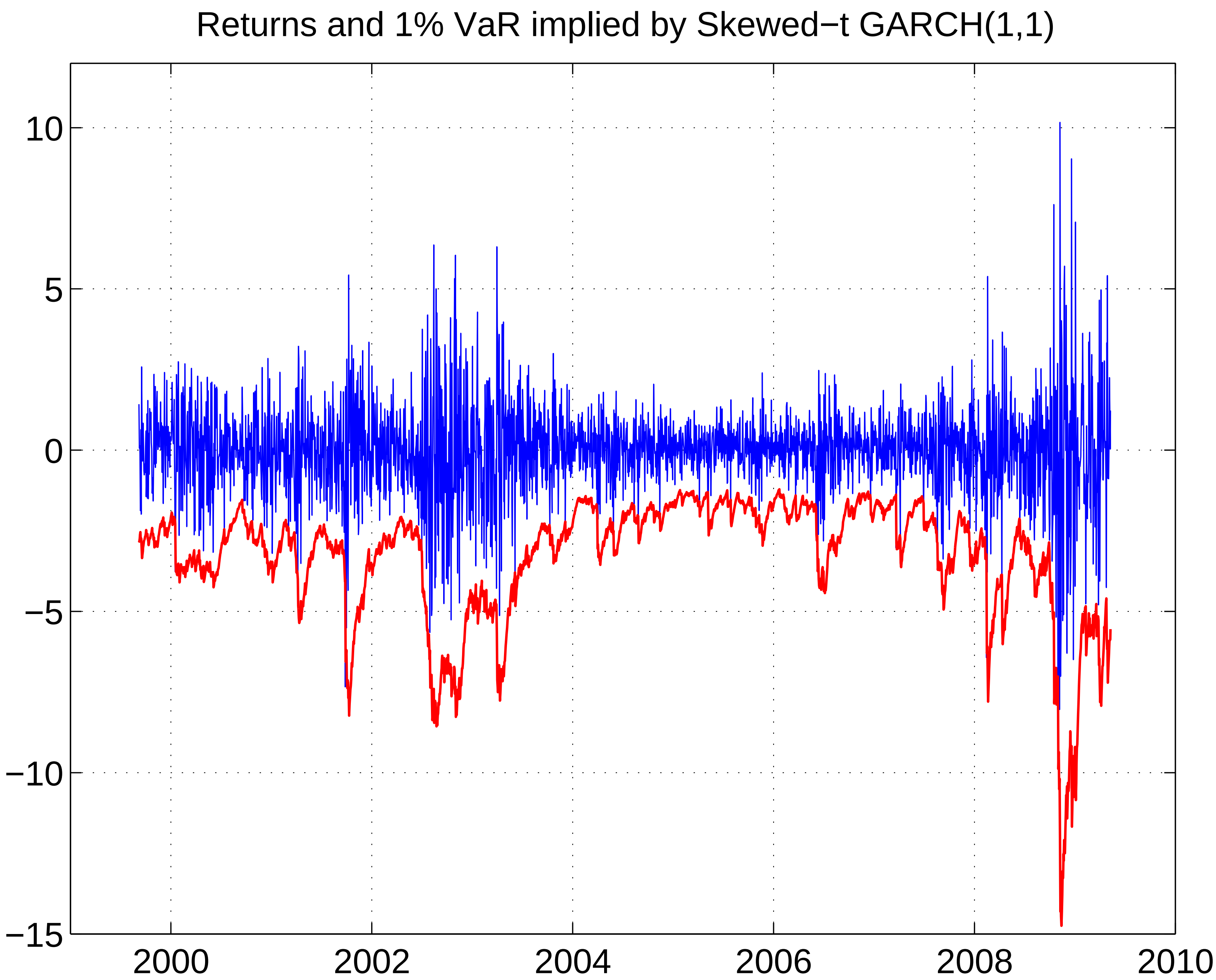

Returns and 1% VaR implied by Skewed−t GARCH(1,1)
10
5
0
−5
−10
−15
2000
2002
2004
2006
2008
2010

Table 10: GARCH(1,1) Value–at–Risk measures, reported is $100 \times \widehat{\xi}$

skewed t GARCH(1,1)

Series	$\xi = 0.001$	0.0025	0.005	0.01	0.025	0.05	0.1
CAC 40	0.16	0.32	0.40	0.85	2.74	5.56	10.93
DAX	0.12	0.24	0.40	0.89	2.58	6.25^{***}	11.61^{***}
FTSE	0.24^{*}	0.24	0.65	1.29	3.15^{**}	5.93^{**}	10.12

Student's t GARCH(1,1)

Series	0.001	0.0025	0.005	0.01	0.025	0.05	0.1
CAC 40	0.20	0.36	0.56	1.17	3.75^{***}	6.49^{***}	11.45^{**}
DAX	0.12	0.28	0.65	1.13	3.43^{***}	7.10^{***}	12.54^{***}
FTSE	0.24^{*}	0.65^{***}	0.85^{**}	1.61^{***}	3.83^{***}	6.57^{***}	11.41^{**}

Asterisks *, **, and *** indicate significance at the 10%, 5% and 1% levels, respectively, based on the test (61).

- Summarizing, both conditional skewness and kurtosis may be important and can considerably improve conditional predictive densities.

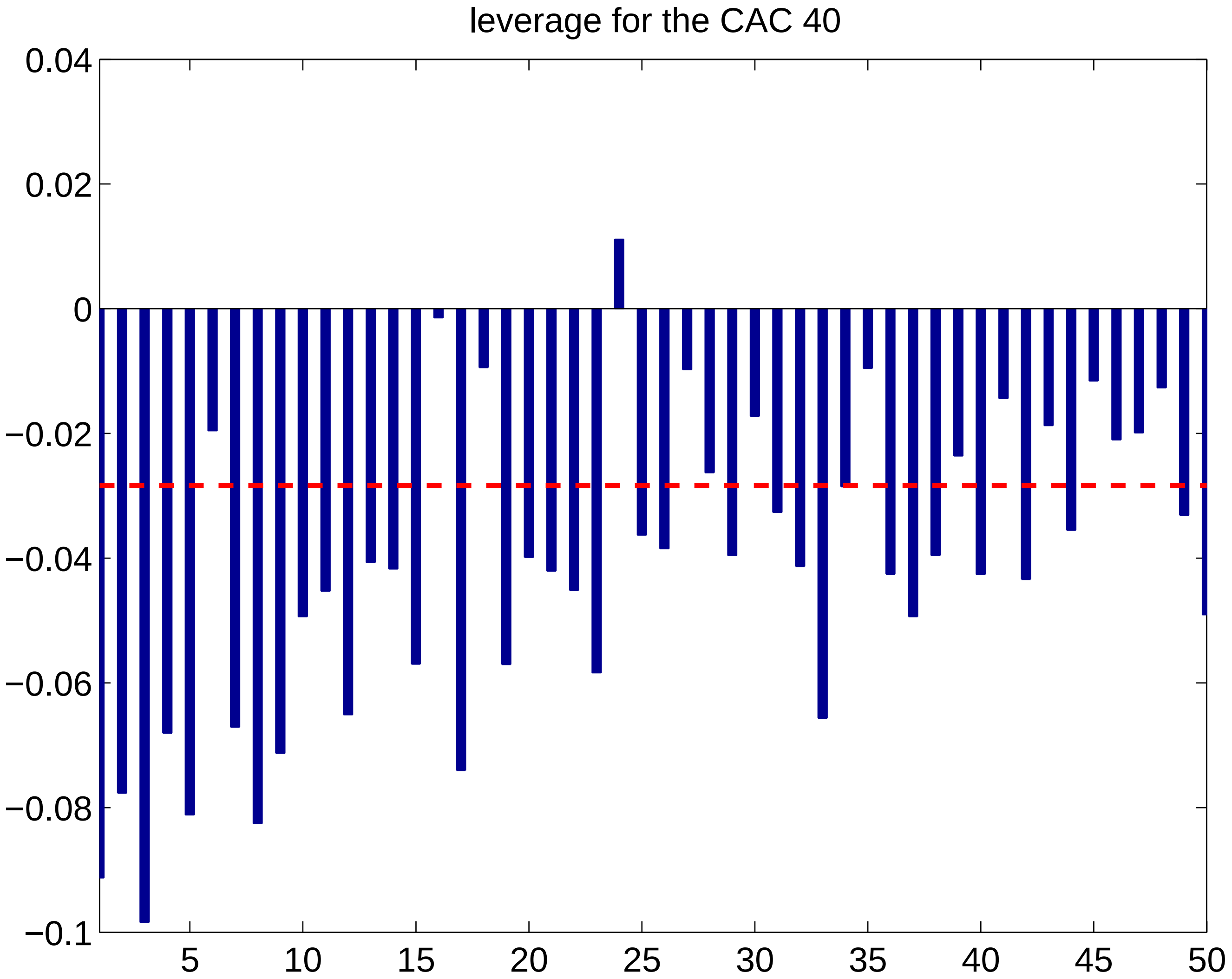
Asymmetric GARCH Models

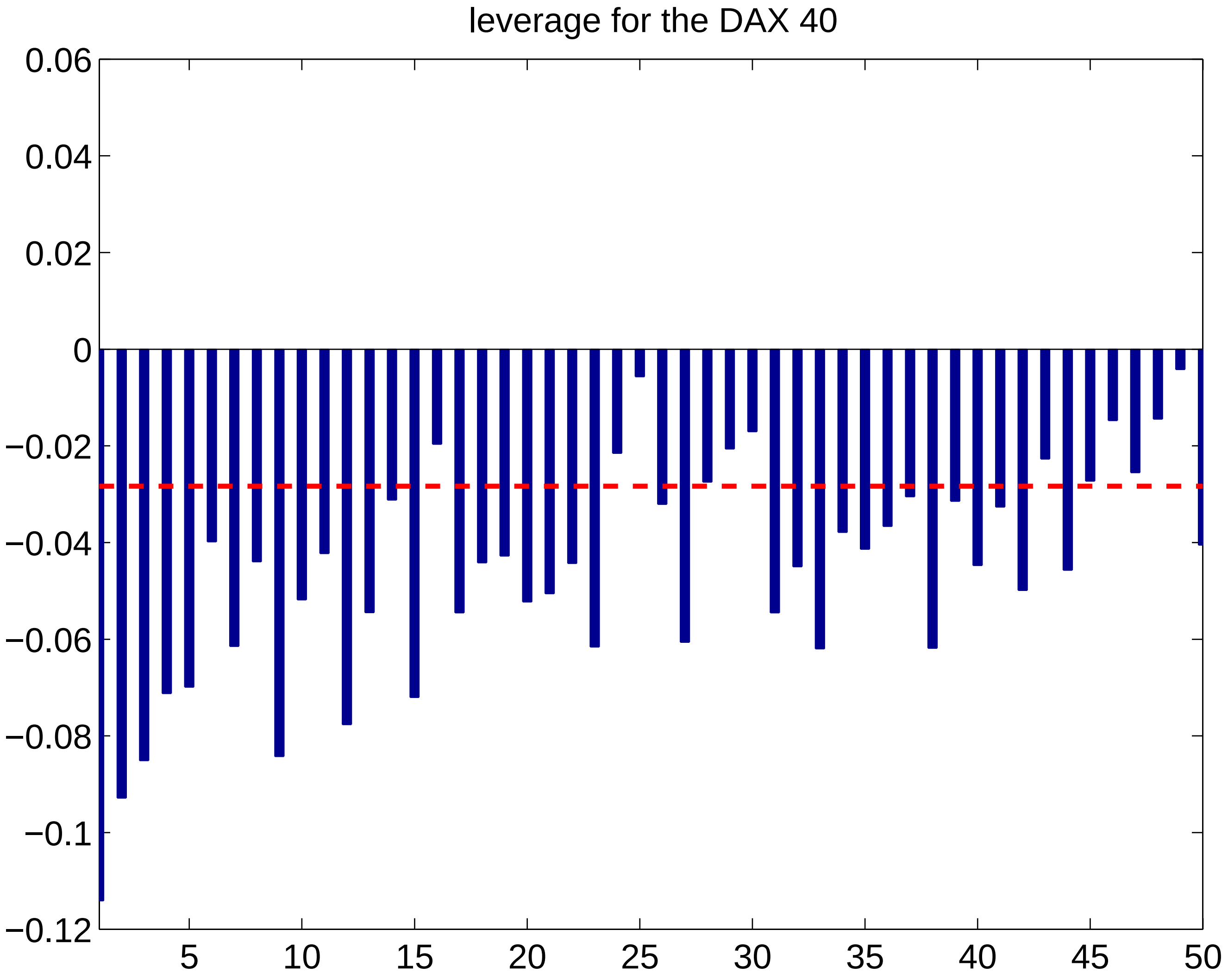
- The basic GARCH model considered so far assumes that the conditional variance σ_t^2 depends only on the *magnitude* and not on the *sign* of past shocks.

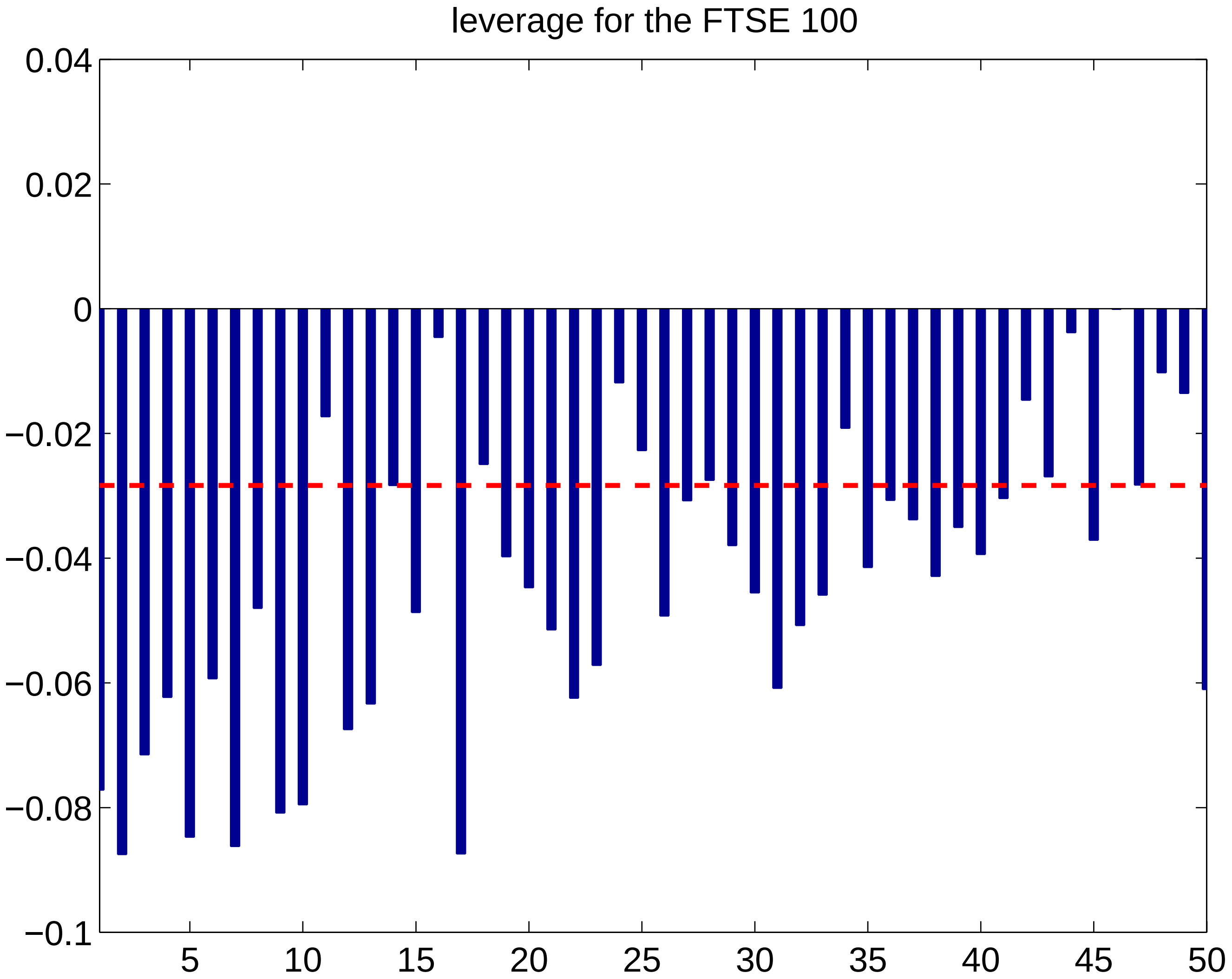
- However, stock market variance tends to react more strongly to bad news than to good news, which is often referred to as the *leverage effect*.

- To illustrate, we may define the leverage effect at lag τ as

$$L(\tau) = \text{Corr}(\epsilon_{t-\tau}, |\epsilon_t|). \tag{64}$$


leverage for the CAC 40
0.04
0.02
0
−0.02
−0.04
−0.06
−0.08
−0.1
5
10
15
20
25
30
35
40
45
50


leverage for the DAX 40
0.06
0.04
0.02
0
−0.02
−0.04
−0.06
−0.08
−0.1
−0.12
5
10
15
20
25
30
35
40
45
50



leverage for the FTSE 100
0.04
0.02
0
−0.02
−0.04
−0.06
−0.08
−0.1
5
10
15
20
25
30
35
40
45
50

Asymmetric GARCH Models I

- The first model that has been put forward is the *Asymmetric GARCH* (AGARCH) of Engle (1990), which specifies the conditional variance as

$$\sigma_t^2 = \omega + \alpha(\epsilon_{t-1} - \theta)^2 + \beta\sigma_{t-1}^2 \quad (65)$$

$$= \omega + \alpha\theta^2 + \alpha\epsilon_{t-1}^2 - 2\alpha\theta\epsilon_{t-1} + \beta\sigma_{t-1}^2. \quad (66)$$

- In model (65), the conditional variance, as a function of ϵ_{t-1}, has its minimum at θ rather than at zero.

- Thus, if $\theta > 0$, negative shocks will have a greater impact on the conditional variance than positive shocks of the same magnitude.

- (66) shows that, if $\alpha + \beta < 1$, the unconditional variance of this process is

$$\mathsf{E}(\sigma_t^2) = \frac{\omega + \alpha\theta^2}{1 - \alpha - \beta}. \quad (67)$$

Asymmetric GARCH Models II

- The asymmetric GARCH model proposed by Glosten, Jagannathan and Runkle (1993), referred to as *GJR–GARCH*, models the conditional variance as
$$\sigma_t^2 = \omega + (\alpha + \theta S_{t-1})\epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2,$$
where
$$S_{t-1} = \begin{cases} 1 & \text{if } \epsilon_{t-1} < 0 \\ 0 & \text{if } \epsilon_{t-1} \geq 0 \end{cases}$$

- Clearly $\theta > 0$ implies that the change in the next period's variance is negatively correlated with today's return.

- If the innovation density is symmetric (e.g., normal or Student's t), the unconditional variance is
$$\mathsf{E}(\sigma_t^2) = \frac{\omega}{1 - \alpha - \theta/2 - \beta}.$$

News Impact Curve

- To analyze the asymmetric response of the variance in different GARCH specifications, Engle and Ng (1993) defined the new impact curve (NIC).

- This is defined as the functional relationship

$$\sigma_t^2 = \sigma_t^2(\epsilon_{t-1}),$$

  with all lagged variances evaluated at their unconditional values.

- For example, for the standard *symmetric* GARCH(1,1) model, we have

$$\sigma_t^2(\epsilon_{t-1}) = A + \alpha\epsilon_{t-1}^2,$$

  where

$$A = \omega + \beta\sigma^2, \quad \sigma^2 = \frac{\omega}{1-\alpha-\beta}.$$

- This is a symmetric function of ϵ_{t-1}.

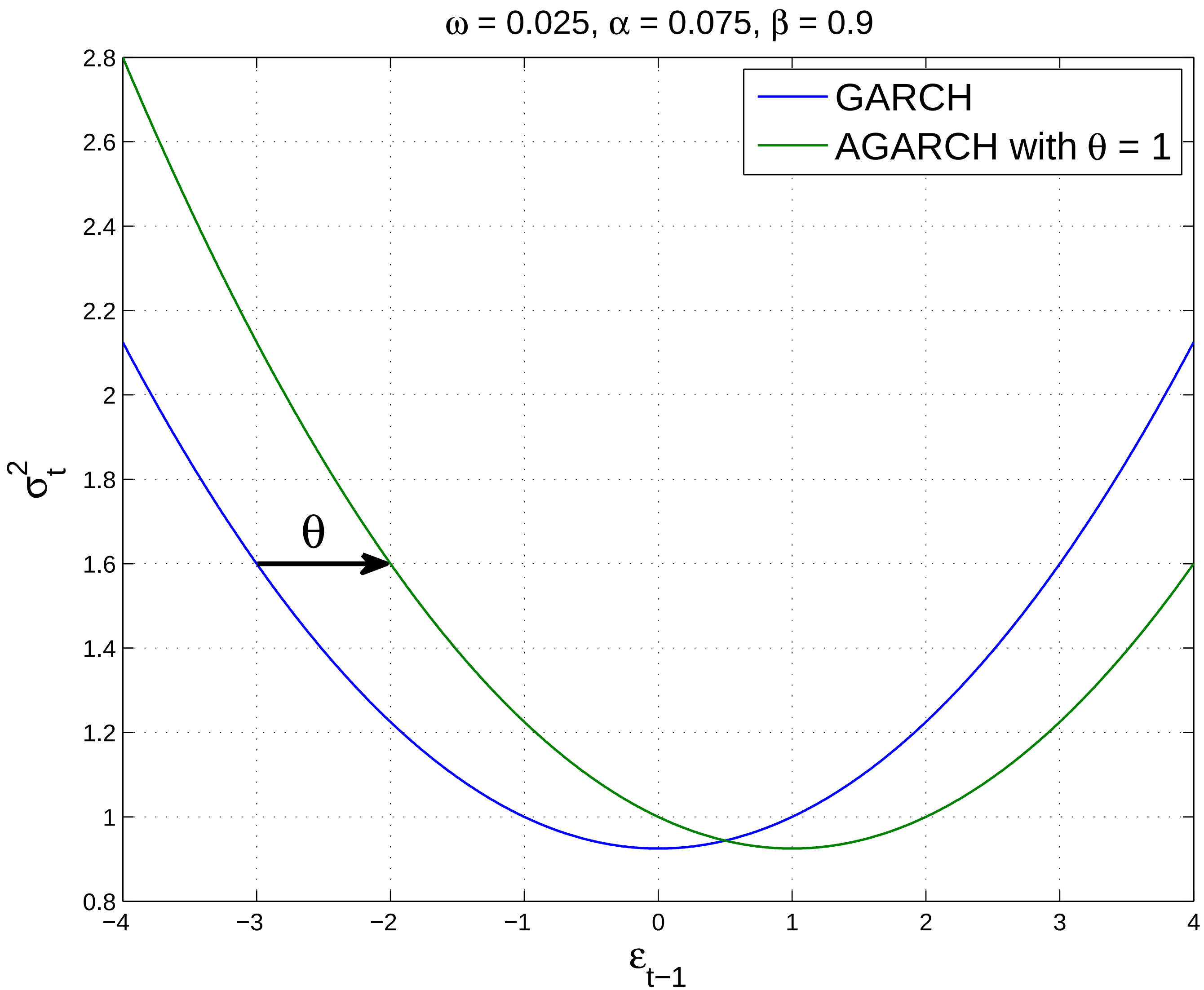
- Asymmetries may be introduced in various ways: Compared to the standard GARCH, we can change either the position of the slope of the NIC (or both).

- For example, the AGARCH captures asymmetry by allowing its NIC to be centered at a positive ϵ_{t-1}, since

$$\sigma_t^2(\epsilon_{t-1}) = A + \alpha(\epsilon_{t-1} - \theta)^2,$$

  where

$$A = \omega + \beta\sigma^2, \quad \sigma^2 = \frac{\omega + \alpha\theta^2}{1 - \alpha - \beta}.$$


ω = 0.025, α = 0.075, β = 0.9
GARCH
AGARCH with θ = 1
θ
2.8
2.6
2.4
2.2
2
1.8
1.6
1.4
1.2
1
0.8
−4
−3
−2
−1
0
1
2
3
4
σ²_t
ε_t−1

- The GJR captures the asymmetry in the impact of news on volatility via a steeper *slope* for negative than for positive shocks, i.e.,

$$\sigma_t^2(\epsilon_{t-1}) = A + \begin{cases} (\alpha + \theta)\epsilon_{t-1}^2 & \text{if } \epsilon_{t-1} < 0 \\ \alpha\epsilon_{t-1}^2 & \text{if } \epsilon_{t-1} \geq 0, \end{cases}$$

  but the NIC of the GJR is still centered at zero, i.e., $\sigma_t^2(\epsilon_{t-1})$ is minimized for $\epsilon_{t-1} = 0$.

- There exist further variants of asymmetric GARCH specifications, e.g., the popular EGARCH (exponential GARCH).

- The estimates reported on the following pages are based on normal innovations; clearly nonnormal distributions allowing for fat tails and asymmetries would be considered in practice.

Table 11: Asymmetric GARCH(1,1) estimates for various stock return series, January 1990 to October 2009

AGARCH (Gaussian)

Series	$\widehat{\omega}$	$\widehat{\alpha}$	$\widehat{\beta}$	$\widehat{\theta}$
CAC 40	0.0000	0.0621	0.9187	0.7361
	(0.0073)	(0.0069)	(0.0084)	(0.0954)
DAX	0.0087	0.0709	0.9081	0.6524
	(0.0069)	(0.0073)	(0.0088)	(0.0829)
FTSE	0.0000	0.0673	0.9189	0.4693
	(0.0036)	(0.0071)	(0.0079)	(0.0664)

GJR–GARCH (Gaussian)

Series	$\widehat{\omega}$	$\widehat{\alpha}$	$\widehat{\beta}$	$\widehat{\theta}$
CAC 40	0.0297	0.0157	0.9184	0.0959
	(0.0050)	(0.0067)	(0.0086)	(0.0109)
DAX	0.0364	0.0220	0.9042	0.1049
	(0.0053)	(0.0072)	(0.0093)	(0.0126)
FTSE	0.0119	0.0187	0.9227	0.0943
	(0.0021)	(0.0064)	(0.0073)	(0.0104)

Table 12: Maximized log–likelihood values

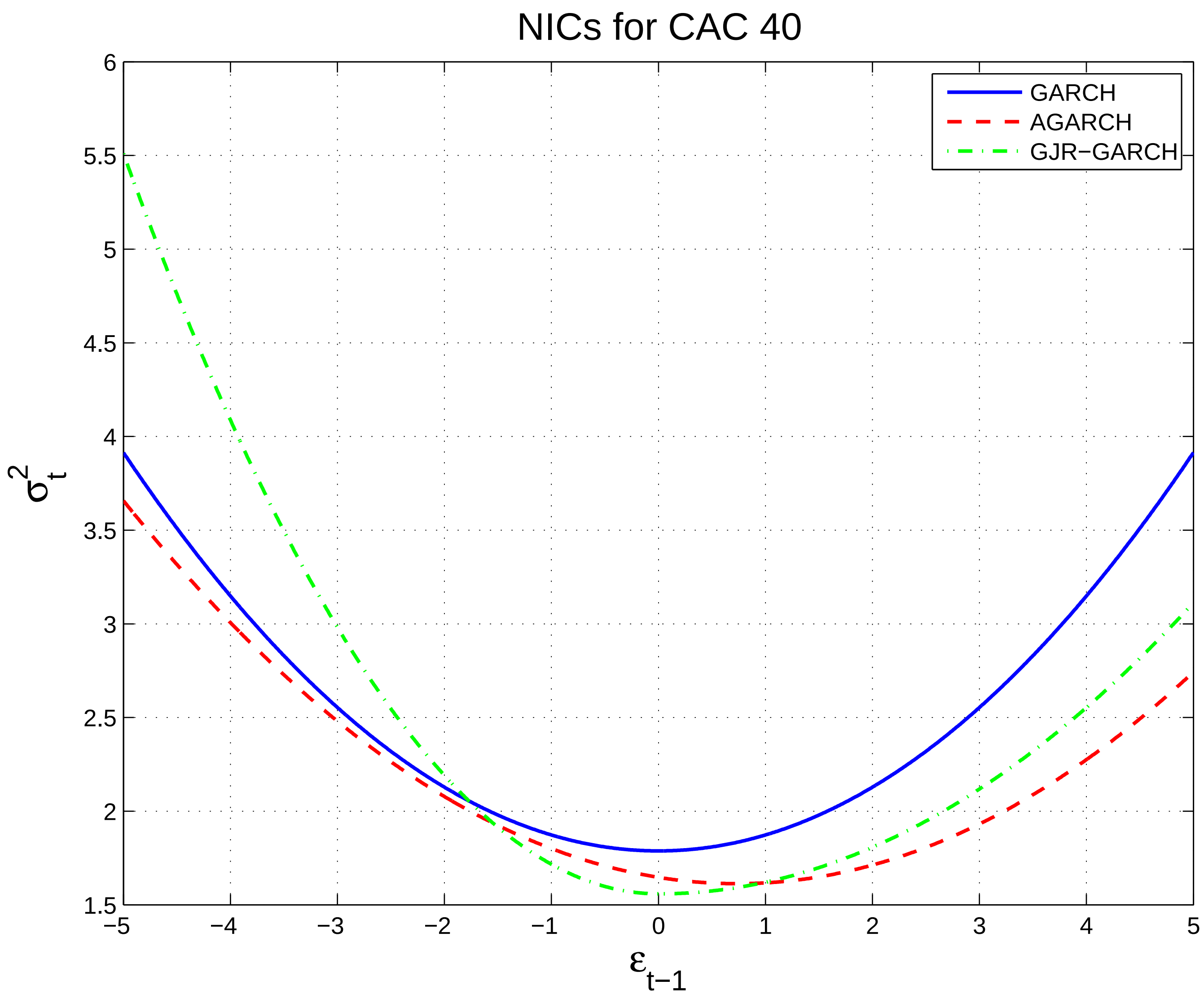
	CAC 40	DAX	FTSE
GARCH	–8088.5	–8180.9	–6798.8
AGARCH	–8045.0	–8141.8	–6761.2
GJR–GARCH	–8043.8	–8138.5	–6755.2

Differences in log–likelihood

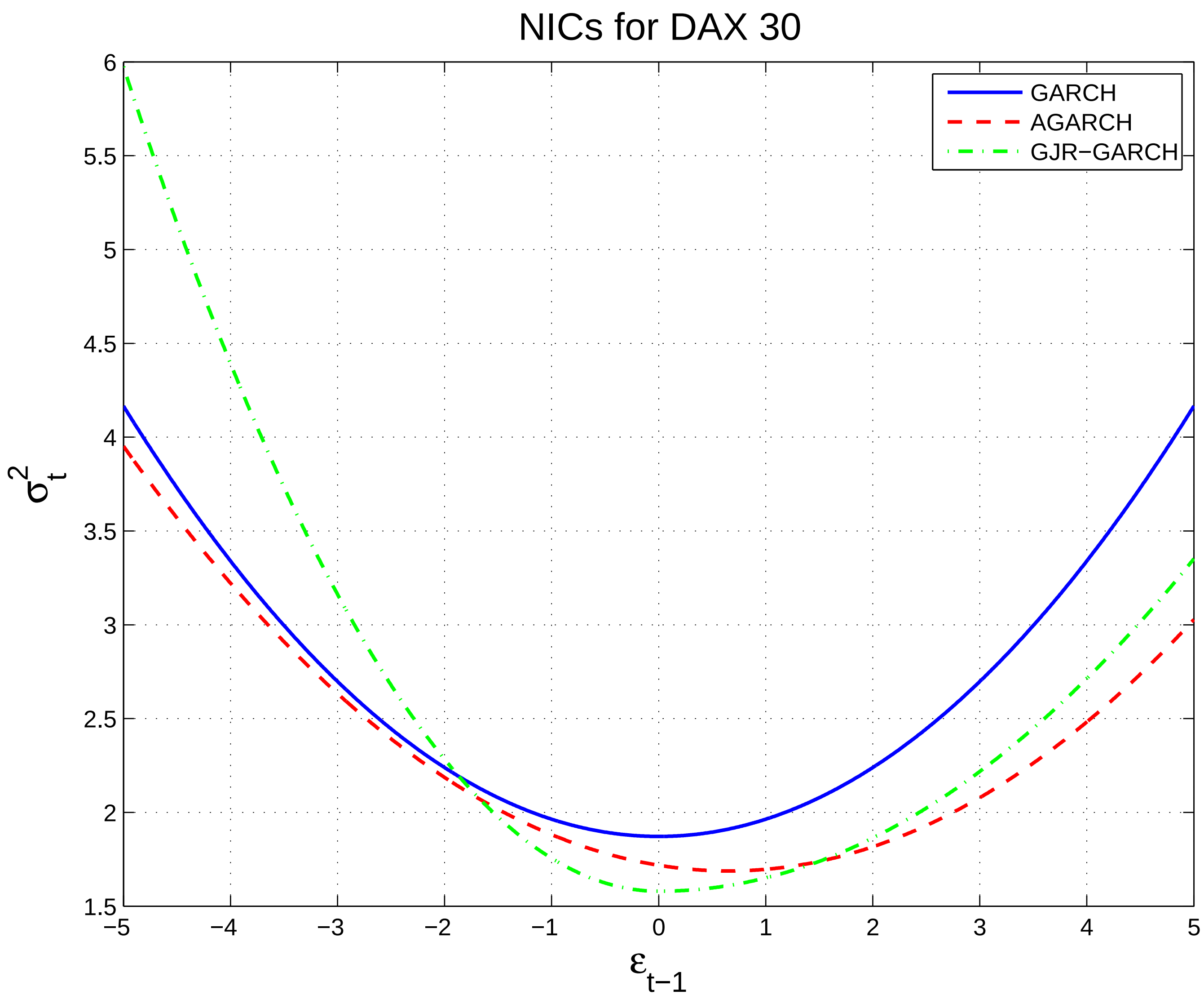
AGARCH – GARCH	43.5299	39.0356	37.6334
GJR – GARCH	44.6940	42.3483	43.6754

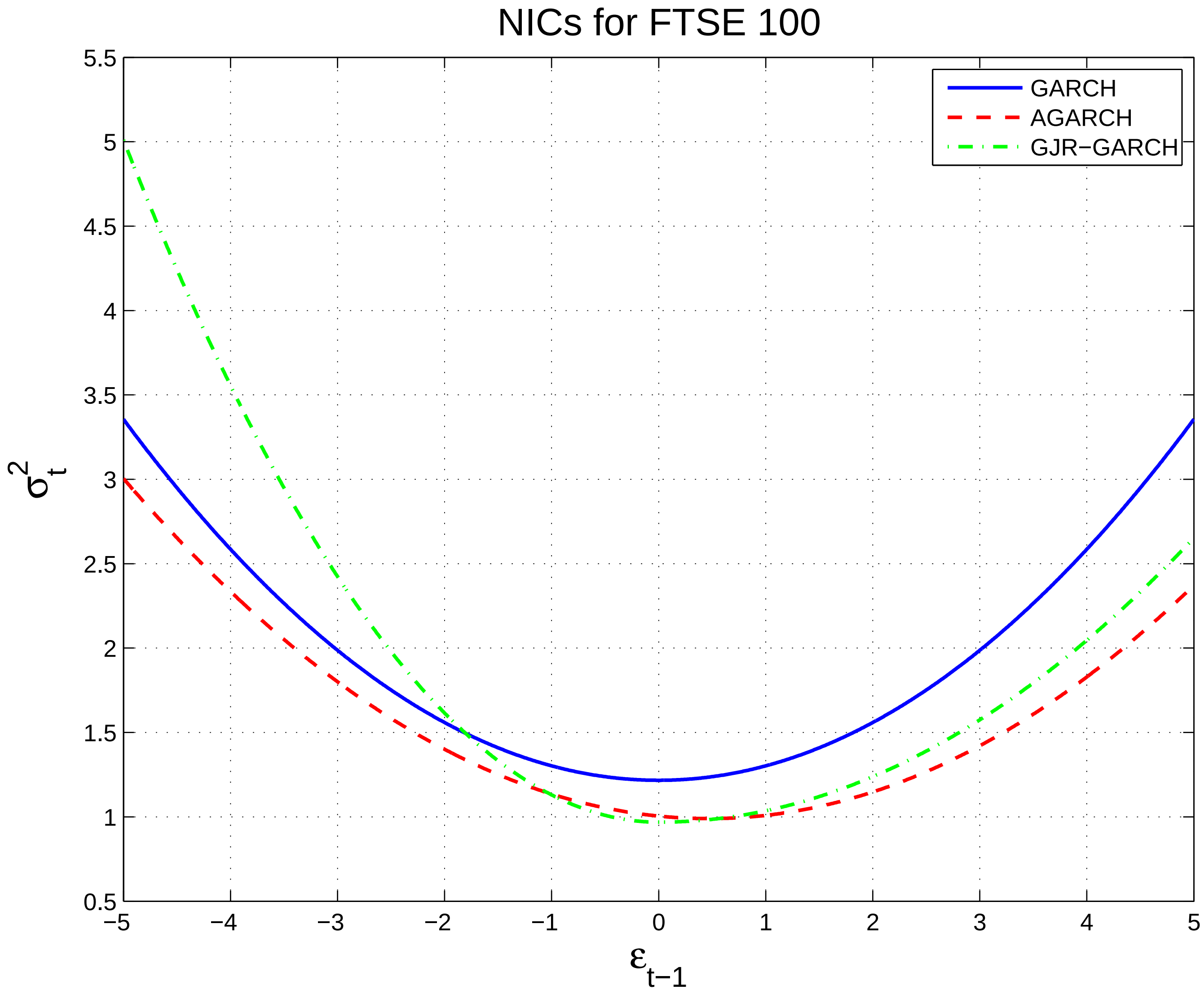
Table 13: Unconditional variances, $\mathsf{E}(\sigma_t^2)$

	CAC 40	DAX	FTSE
data	2.0016	2.2133	1.3231
GARCH	1.9542	2.0610	1.3306
AGARCH	1.7559	1.8493	1.0772
GJR–GARCH	1.6649	1.7070	1.0370



NICs for CAC 40
GARCH
AGARCH
GJR−GARCH
σ_t^2
ε_{t-1}
6
5.5
5
4.5
4
3.5
3
2.5
2
1.5
−5
−4
−3
−2
−1
0
1
2
3
4
5


NICs for DAX 30
σ_t^2
ε_{t-1}
GARCH
AGARCH
GJR−GARCH
6
5.5
5
4.5
4
3.5
3
2.5
2
1.5
−5
−4
−3
−2
−1
0
1
2
3
4
5


NICs for FTSE 100
GARCH
AGARCH
GJR−GARCH
σ_t^2
ε_t−1
5.5
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
−5
−4
−3
−2
−1
0
1
2
3
4
5

ARCH–M

- In the finance literature, a link is often made between the expected return and the risk of an asset.

- Investors are willing to hold risky assets only if their expected return compensate for the risk.

- A model that incorporates this link is the GARCH–in–mean or GARCH–M model, which can be written as

$$r_t = c + \delta g(\sigma_t^2) + \epsilon_t,$$

  where ϵ_t is a GARCH error process, and g is a known function such as $g(\sigma_t^2) = \sigma_t^2$, $g(\sigma_t^2) = \sigma_t$, or $g(\sigma_t^2) = \log(\sigma_t^2)$.

- If $\delta > 0$ and g is monotonically increasing, then the term $\delta g(\sigma_t^2)$ can be interpreted as a *risk premium* that increases expected returns if conditional volatility σ_t^2 is high.

- In practice $g(\sigma_t^2) = \sigma_t$ appears to be the preferred specification.