EUROPEAN JOURNAL OF PURE AND APPLIED MATHEMATICS
Vol. 3, No. 3, 2010, 443-477
ISSN 1307-5543 – www.ejpam.com

SPECIAL ISSUE ON
GRANGER ECONOMETRICS AND STATISTICAL MODELING
DEDICATED TO THE MEMORY OF PROF. SIR CLIVE W.J. GRANGER

Stylized Facts of Financial Time Series and Three Popular Models of Volatility

Hans Malmsten[1] and Timo Teräsvirta[2*]

[1] *Länsförsäkringar, Stockholm*
[2] *CREATES, Aarhus University, Building 1322, DK-8000 Aarhus C, Denmark*

Abstract. Properties of three well-known and frequently applied first-order models for modelling and forecasting volatility in daily or weekly financial series such as stock and exchange rate returns are considered. These are the standard Generalized Autoregressive Conditional Heteroskedasticity (GARCH), the Exponential GARCH and the Autoregressive Stochastic Volatility model. The focus is on finding out how well these models are able to reproduce characteristic features of such series, also called stylized facts. These include high kurtosis and a rather low-starting and slowly decaying autocorrelation function of the squared or absolute-valued observations. Another stylized fact is that the autocorrelations of absolute-valued returns raised to a positive power are maximized when this power equals unity. Not unexpectedly, a conclusion that emerges from these considerations, largely based on results on the moment structure of these models, is that none of the models dominates the others when it comes to reproducing stylized facts in typical financial time series.

2000 Mathematics Subject Classifications: 62M10, 91B84

Key Words and Phrases: autocorrelation of squared residuals, autoregressive conditional heteroskedasticity, conditional variance, kurtosis, stochastic volatility, volatility of stock returns

1. Introduction

Modelling volatility of financial series such as stock returns has become common practice, as the demand for volatility forecasts has increased. Various types of models such as models of autoregressive conditional heteroskedasticity and stochastic volatility models have been applied for the purpose. A practitioner can thus choose between a variety of models. A popular way of comparing volatility models has been to estimate a number of models by

*Corresponding author.

Email address: tterasvirta@creates.au.dk (T. Teräsvirta)

maximum likelihood and observe which one has the highest log-likelihood value; see [48] for an example. If the models under comparison do not have the same number of parameters, one may want to favour parsimony and apply a suitable model selection criterion, such as AIC or BIC, for the purpose. It is also possible to choose a model after actually applying it to forecasting. [44] provide a survey of papers that contain results of such comparisons.

Another way of comparing models is to submit estimated models to misspecification tests and see how well they pass the tests. This also paves the way for building models within the same family of models. One can extend a failed model by estimating the alternative it has been tested against and subject that model to new misspecification tests. Such tests have been derived for generalized autoregressive conditional heteroskedasticity (GARCH) models; see, for example, [15, 10, 37, 39]. Similar devices for the exponential GARCH (EGARCH) model of [42] who already suggested such tests, are presented in [40]. In addition, nonnested models can be tested against each other. [33] considered testing GARCH against the autoregressive stochastic volatility (ARSV) model and [35] suggested the simulated likelihood ratio test for choosing between GARCH and EGARCH: for other approaches see [15] and [38]. The pseudo-score test of [9] can be applied to this problem as well. Small sample properties of some of the available tests for that testing problem are considered in [40]. It should be noted, however, that testing two models against each other does not necessary lead to a unique choice of a model. Neither model may be rejected against the other or both may be rejected against each other. For a discussion of conceptual differences between the model selection and testing approaches, see [21].

The purpose of this paper is to compare volatility models from a rather different angle. Financial time series of sufficiently high frequency such as daily or weekly or even intradaily stock or exchange rate return series seem to share a number of characteristic features, sometimes called stylized facts. [20] and [22], among others, pointed out such features and investigated their presence in financial time series. Given a set of characteristic features or stylized facts, one may ask the following question: "Have popular volatility models been parameterized in such a way that they can accommodate and explain the most common stylized facts visible in the data?" Models for which the answer is positive may be viewed as suitable for practical use. The other parameterizations may be regarded as less useful in practice.

There exists some work towards answering this question. [51] considered the ability of the GARCH model to reproduce series with high kurtosis and, at the same time, positive but low and slowly decreasing autocorrelations of squared observations. [36] discussed this stylized fact in connection with the ARSV model, whereas [2] focussed on the ARSV model based on the normal inverse Gaussian distribution. [8] compared the ARSV model and the GARCH model using the kurtosis-autocorrelation relationship as their benchmark. [3] also compared GARCH and ARSV models. The work of [45] on the hidden Markov model for the variance may also be mentioned in this context. Furthermore, [55] considered stylized facts similar to the ones discussed in this paper in the context of hyperbolic diffusions.

Answering the question by using the approach of this paper is only possible in the case of rather simple models. On the other hand, a vast majority of popular models such as GARCH, EGARCH and ARSV models used in applications are first-order models. Higher-order models, although theoretically well-defined, are rather seldom used in practice. This suggests that restricting the considerations to simple parameterizations does not render the results useless.

This paper may be viewed as an extension to [51] and has the following contents. The stylized facts are defined in Section 2 and the models are discussed in Section 3. Section 4 considers the kurtosis-autocorrelation relationship. In Section 5, a stylized fact called the Taylor effect is discussed. In Section 6 the kurtosis-autocorrelation relationship is reconsidered using confidence regions. A stylized fact that cannot be reproduced by the models under consideration but has generated plenty of discussion is briefly mentioned in Section 7. Section 8 contains conclusions.

2. Stylized Facts

The stylized facts to be discussed in this paper are illuminated by Figure 1. The first panel depicts the return series of the S&P 500 stock index (daily first differences r_t of logarithms of the index; 19261 observations) from 3 January 1928 to 24 April 2001. The marginal distribution of r_t appears leptokurtic and a number of volatility clusters are clearly visible. The volatility models considered in this study are designed for parameterizing this type of variation. The second panel shows the autocorrelation function of $|r_t|^m$, $m = 0.25, 0.5, 0.75, 1$ and the third one the corresponding function for $m = 1, 1.25, 1.5, 1.75, 2$, for the first 500 lags. It is seen that the first autocorrelations have positive but relatively small values and that the autocorrelations decay slowly. A similar figure can be found in [12], but here the time series has been extended to cover ten more years from 1992 to 2001.

The first stylized fact illustrated by Figure 1 and typical of a large amount of return series is the combination of relatively high kurtosis and rather low autocorrelations of $|r_t|^m$. In the case of the standard GARCH model, we restrict ourselves to inspect the combination of kurtosis and the autocorrelations of r_t^2 because in that case, an analytic expression for the autocorrelation function is available. The second stylized fact to be considered is the fact that the autocorrelations as a function of m tend to peak for $m = 1$. This is the so-called Taylor effect that has been found in a large number of financial time series; see [20] and [22]. In the GARCH framework, this stylized fact can only be investigated using analytic expressions when the GARCH model is the so-called absolute-value GARCH (AVGARCH) model and $m = 1$ or $m = 2$. This is because no analytical expressions for ρ $(|r_t|^m, |r_{t-j}|^m)$ exist when $m < 2$ and the model is the standard GARCH model. For the AVGARCH model, they are available for both $m = 1$ and $m = 2$ but not for non-integer values of m.

Yet another fact discernible in Figure 1 is that the decay rate of the autocorrelations is very slow, apparently slower than the exponential rate. This prompted the introduction of the fractionally integrated GARCH (FIGARCH) model; see [4]. Other approaches to this problem include the locally stationary ARCH model by [11], and the globally nonstationary GARCH model of which there exist different versions, see [56, 16, 1, 5]. In this paper, this slow decay is not included among the stylized facts under consideration. In this work we concentrate on weakly stationary GARCH models and also exclude the FIGARCH model. To illustrate the reason for this exclusion, we split the S&P 500 return series into 20 subseries of 980 observations each and estimate the autocorrelations ρ $(|r_t|, |r_{t-j}|), j = 1, ..., 500$, for these subseries. The lowest panel of Figure 1 contains these autocorrelations for the whole series and the mean of the corresponding autocorrelations of the 20 subseries together with the plus/minus one standard deviation band. It is seen that the decay of autocorrelations in the subseries on the average is substantially faster than in the original series and roughly exponential. This lack of self-similarity in autocorrelations can be taken as evidence against the FIGARCH model in this particular case, but that is beside the point. (For more discussion, see [41]). We merely want to argue that the very slow decay rate of the autocorrelations of $|r_t|$ or r_t^2 may not necessarily be a feature typical of series with a couple of thousand observations. Because such series are most often modelled by one of the standard models of interest in this study, we do not consider a very slow decay rate of autocorrelations a stylized fact in our discussion.

3. The Models and their Fourth-moment Structure

3.1. GARCH Model

Suppose an error term or an observable variable can be decomposed as follows:

$$\varepsilon_t = z_t h_t^{1/2} \tag{1}$$

where $\{z_t\}$ is a sequence of independent identically distributed random variables with zero mean and finite variance. Furthermore, assume that

$$h_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j \varepsilon_{t-j}^2 + \sum_{j=1}^{p} \beta_j h_{t-j}. \tag{2}$$

Equations (1) and (2) define the standard GARCH(p,q) model of [7]. Parameter restrictions are required to ensure positiveness of the conditional variance h_t in (2). Assuming $\alpha_j \geqslant 0$, $j = 1,...,q$, and $\beta_j \geqslant 0$, $j = 1,...,p$, is sufficient for this. The GARCH model has since its introduction been generalized in various directions, see [52] for a recent survey. Both necessary and sufficient conditions were derived by [43]. In this paper we shall concentrate on (1) with (2) assuming $p = q = 1$. This is done for two reasons. First, the GARCH(1,1) model is by far the most frequently applied GARCH specification. Second, we want to keep our considerations simple.

The GARCH(1,1) model is covariance stationary if

$$\alpha_1 \nu_2 + \beta_1 < 1 \tag{3}$$

where $\nu_2 = \mathsf{E} z_t^2 < \infty$. For the discussion of stylized facts we need moment condition and fourth moments of $\{\varepsilon_t\}$. Assuming $\nu_4 = \mathsf{E} z_t^4 < \infty$, the unconditional fourth moment for the GARCH(1,1) model exists if and only if

$$\alpha_1^2 \nu_4 + 2\alpha_1 \beta_1 \nu_2 + \beta_1^2 < 1. \tag{4}$$

Under (4) the kurtosis of ε_t equals

$$\kappa_4 = \frac{\kappa_4(z_t)\{1 - (\alpha_1 \nu_2 + \beta_1)^2\}}{1 - (\alpha_1^2 \nu_4 + 2\alpha_1 \beta_1 \nu_2 + \beta_1^2)} \tag{5}$$

where $\kappa_4(z_t) = \nu_4/\nu_2^2$ is the kurtosis of z_t. Assuming normality, one obtains the following well-known result:

$$\kappa_4 = 3\frac{1-(\alpha_1+\beta_1)^2}{1-(3\alpha_1^2+2\alpha_1\beta_1+\beta_1^2)} > 3. \tag{6}$$

Furthermore, when (4) holds, the autocorrelation function of $\{\varepsilon_t^2\}$ is defined as follows:

$$\rho_n = (\alpha_1\nu_2+\beta_1)^{n-1}\frac{\alpha_1\nu_2(1-\beta_1^2-\beta_1\alpha_1\nu_2)}{1-\beta_1^2-2\beta_1\alpha_1\nu_2}\ n \geq 1. \tag{7}$$

The autocorrelation function of $\{\varepsilon_t^2\}$ is dominated by an exponential decay from the first lag with decay rate $\alpha_1\nu_2+\beta_1$. Setting $\nu_2 = 1$ and $\nu_4 = 3$ (normality) in (7) gives the result in [6]. Note that the existence of the autocorrelation function does depend on the existence of ν_4 although (7) is not a function of ν_4. The necessary and sufficient conditions for the existence of the unconditional fourth moments of the GARCH(*p,q*) process and the expressions (5) and (7) are special cases of results in [24].

3.2. EGARCH Model

[42] who introduced the EGARCH model listed three drawbacks with the GARCH models. First, there is the lack of asymmetry in the response to shocks. Second, parameter restrictions have to be imposed on the GARCH model to ensure positivity of the conditional variance. Finally, persistence is an ambiguous concept. Consider (1) with

$$\ln h_t = \alpha_0 + \sum_{j=1}^{q}\{\phi_j z_{t-j} + \psi_j(\left|z_{t-j}\right| - \mathsf{E}\left|z_{t-j}\right|)\} + \sum_{j=1}^{p}\beta_j \ln h_{t-j} \tag{8}$$

which defines the EGARCH(*p,q*) model of [42]. It is seen from (8) that no parameter restrictions are necessary to ensure positivity of h_t. The fourth-moment structure of the EGARCH(*p,q*) model has been worked out in [28] and [32]. As in the GARCH case, the first-order model is the most popular EGARCH model. The term $\psi(\left|z_{t-1}\right| - \mathsf{E}\left|z_{t-1}\right|)$ represents a magnitude effect in the spirit of the GARCH(1,1) model. The term ϕz_t represents the asymmetry effect. [42] derived existence conditions for moments of the EGARCH(1,1) model. Setting $\beta = \beta_1$, they can be summarized by saying that if the error process $\{z_t\}$ has all moments then all moments for the EGARCH(1,1) process exist if and only if

$$\left|\beta\right| < 1. \tag{9}$$

For example, if $\{z_t\}$ is standard normal then the restriction (9) is both necessary and sufficient for the existence of all moments. This is different from the GARCH model. For that model, the moment conditions become more and more restrictive when the order of the moment increases.

Another difference between the GARCH models and the EGARCH model is that for the latter analytical expressions exist for all moments of $\left|\varepsilon_t\right|^{2m}$, $m > 0$. They can be found in

[27]; see also [42]. If (9) holds, then the kurtosis of ε_t, assuming z_t ∼nid(0,1), is given by

$$\begin{aligned}
\kappa_4 &= 3\exp\{\frac{(\psi+\phi)^2}{1-\beta^2}\}\prod_{i=1}^{\infty}\frac{\Phi(2\beta^{i-1}(\psi+\phi))+\exp\{-8\beta^{2(i-1)}\psi\phi\}\Phi(2\beta^{i-1}(\psi-\phi))}{[\Phi(\beta^{i-1}(\psi+\phi))+\exp\{-2\beta^{2(i-1)}\psi\phi\}\Phi(\beta^{i-1}(\psi-\phi))]^2} \\
&> 3 \qquad (10)
\end{aligned}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. The expression contains infinite products, and care is therefore required in computing them (selecting the number of terms in the product). Setting $\psi = 0$ in (10) yields a simple formula

$$\kappa_4 = 3\exp\{\phi^2(1-\beta^2)^{-1}\} > 3. \qquad (11)$$

If (9) holds, the autocorrelation function for $\left|\varepsilon_t\right|^{2m}$, with z_t ∼nid(0,1), has the form

$$\begin{aligned}
&\rho_n(m) \\
&= \frac{\frac{\Gamma(2m+1)}{2^{m+1/2}\Gamma(m+1/2)}\exp\{\frac{m^2(\psi+\phi)^2(\beta^{2(n-1)}(\beta^2-1)/4+\beta^n)}{1-\beta^2}\}D(\cdot)\prod_{i=1}^{n-1}\Phi_{1i}\prod_{i=1}^{\infty}\Phi_{2i}-\prod_{i=1}^{\infty}\Phi_{1i}^2}{\frac{\pi^{1/2}\Gamma(2m+1/2)}{(\Gamma(m+1/2))^2}\exp\{\frac{m^2(\psi+\phi)^2}{1-\beta^2}\}\prod_{i=1}^{\infty}\Phi_{3i}-\prod_{i=1}^{\infty}\Phi_{1i}^2} \\
&\quad n \geq 1 \qquad (12)
\end{aligned}$$

where

$$\begin{aligned}
D(\cdot) &= D_{-(2m+1)}[-m\beta^{n-1}(\psi+\phi)]+\exp\{-m^2\beta^{2(n-1)}\psi\phi\} \\
&\quad \times D_{-(2m+1)}[-m\beta^{n-1}(\psi-\phi)] \\
\Phi_{1i} &= \Phi(m\beta^{i-1}(\psi+\phi))+\exp\{-2m^2\beta^{2(i-1)}\psi\phi\}\Phi(m\beta^{i-1}(\psi-\phi)) \\
\Phi_{2i} &= \Phi(m\beta^{i-1}(1+\beta^n)(\psi+\phi))+\exp\{-2m^2\beta^{2(i-1)}(1+\beta^n)^2\psi\phi\} \\
&\quad \times\Phi(m\beta^{i-1}(1+\beta^n)(\psi-\phi))
\end{aligned}$$

and

$$\Phi_{3i} = \Phi(2m\beta^{i-1}(\psi+\phi))+\exp\{-8m^2\beta^{2(i-1)}\psi\phi\}\Phi(2m\beta^{i-1}(\psi-\phi)).$$

Furthermore, $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution and

$$D_{(-p)}[q] = \frac{\exp\{-q^2/4\}}{\Gamma(p)}\int_0^{\infty}x^{p-1}\exp\{-qx-x^2/2\}dx,\ p>0,$$

is the parabolic cylinder function where $\Gamma(\cdot)$ is the Gamma function. If $\phi = 0$ or $\psi = 0$ in the EGARCH(1,1) model the resulting autocorrelation function becomes quite simple; see [27]. The autocorrelation function of the squared observations ($m = 1$), when $\psi = 0$, has the form

$$\rho_n(1) = \frac{(1+\phi^2\beta^{2(n-1)})\exp\{\phi^2\beta^n(1-\beta^2)^{-1}\}-1}{3\exp\{\phi^2(1-\beta^2)^{-1}\}-1}, n \geq 1. \qquad (13)$$

To illustrate the above theory, consider the case $0 < \beta < 1$. The decay of the autocorrelations is controlled by the parameter β. The autocorrelation function of $\{|\varepsilon_t|^{2m}\}$ then has the property that the decay rate is faster than exponential at short lags and approaches β as the lag length increases. For the special case (13) this can be shown analytically, but in the general case it is just a conjecture based on numerical calculations; see the table in [27].

3.3. ARSV Model

The ARSV model offers yet another way of characterizing conditional heteroskedasticity. See [17] and [49] for useful surveys. It bears certain resemblance to the EGARCH model. As with the EGARCH model, defining the dynamic structure using $\ln h_t$ and its lags ensures that h_t is always positive, but the difference to the GARCH model and the EGARCH model is that it does not depend on past observations but on some unobserved latent variable instead. The simplest and most popular ARSV(1) model, [50], is given by

$$\varepsilon_t = \sigma z_t h_t^{1/2}. \tag{14}$$

where σ is a scale parameter. It removes the need for a constant term in the first-order autoregressive process

$$\ln h_{t+1} = \beta \ln h_t + \eta_t. \tag{15}$$

In (15), $\{\eta_t\}$ is a sequence of independent normal distributed random variables with mean zero and a known variance σ_η^2. The error processes $\{z_t\}$ and $\{\eta_t\}$ are assumed to be mutually independent. One motivation for the EGARCH model has been the need to capture the non-symmetric response to the sign of the shock. If z_t and η_t are assumed to be correlated with each other, the ARSV(1) model also allows for asymmetry. The model can be generalized such that $\ln h_t$ follows an ARMA(p,q) process, but in this work we only consider the ARSV(1) model.

As η_t is normally distributed, $\ln h_t$ is also normally distributed. From standard theory we know that all moments of $\ln h_t$ exist if and only if

$$|\beta| < 1 \tag{16}$$

in (15). Thus, if $|\beta| < 1$ and all moments of z_t exist then all moment of ε_t in (14) exist as well, as they do in the EGARCH(1,1) model. If condition (16) is satisfied, the kurtosis of ε_t is given by

$$\kappa_4 = \kappa_4(z_t) \exp\{\sigma_h^2\}, \tag{17}$$

where $\sigma_h^2 = \sigma_\eta^2/(1-\beta^2)$ is the variance of $\ln h_t$. Thus $\kappa_4 > \kappa_4(z_t)$, so that if $z_t \sim$nid(0,1), ε_t is leptokurtic. Formula (17) bears considerable resemblance to (11). In the ARSV(1) model (14) and (15), z_t and η_t are independent. The same is true for z_t and z_{t-1} in the EGARCH(1,1) model. When $\psi_1 = 0$ in the latter model, the moment expressions for the two models therefore look alike.

As in EGARCH models it is possible to derive the autocorrelation function for any $|\varepsilon_t|^{2m}$, $m > 0$, when $\{\varepsilon_t\}$ obeys an ARSV(1) model (14) and (15). When (16) holds, then the autocorrelation function of $\{\left|\varepsilon_t\right|^{2m}\}$ is defined as follows, see [17]:

$$\rho_n(m) = \frac{\exp(m^2\sigma_h^2\beta^n) - 1}{\kappa_m \exp(m^2\sigma_h^2) - 1}, n \geqslant 1, \tag{18}$$

where κ_m equals

$$\kappa_m = E\left|z_t\right|^{4m} / (E\left|z_t\right|^{2m})^2. \tag{19}$$

The autocorrelation function of $\{\left|\varepsilon_t\right|^{2m}\}$ has the property that the decay rate is faster than exponential at short lags and stabilizes to β as the lag length increases, analogously to the EGARCH model. Thus, the decay of the autocorrelations is controlled by β only.

4. Kurtosis-autocorrelation Relationship

4.1. GARCH(1,1) Model

The results in the preceding section make it possible to consider how well the models fits the first stylized fact of financial time series mentioned in Section 2: leptokurtosis and low but rather persistent autocorrelation of the squared observations or errors. Consider GARCH(1,1) model with normal errors and express the autocorrelation function (7) as a function of the kurtosis (5). This yields

$$\rho_n = (\alpha_1 + \beta_1)^{n-1}(\frac{\beta_1(1 - 3\kappa_4^{-1})}{3(1 - \kappa_4^{-1})} + \alpha_1), n \geq 1. \tag{20}$$

Figure 2 illuminates the relationship between the kurtosis κ_4 and the autocorrelation ρ_1. It contains isoquants, curves defined by sets of points for which the sum $\alpha_1 + \beta_1$ has the same value. The kurtosis and the first-order autocorrelation of squared observations are both increasing functions of α_1 when $\alpha_1 + \beta_1$ equals a constant. They all start at $\kappa_4 = 3$ and $\rho_1 = 0$ where $\alpha_1 = 0$ and the GARCH(1,1) model is unidentified (the conditional variance equals unity). For previous examples of similar figures, see [51, 36, 2]. Slightly different contour plots for the GARCH(1,1) model can be found in [3]. It is seen from the present figure that the first-order autocorrelation first increases rapidly as a function of the kurtosis (and α_1) and that the increase gradually slows down. It is also clear that the autocorrelation decreases as a function of $\alpha_1 + \beta_1$ when the kurtosis is held constant. Nevertheless, low autocorrelations cannot exist with high kurtosis.

This figure offers a useful background for studying the observed kurtosis-autocorrelation combinations. Figure 3 contains the same isoquants as the Figure 2, together with kurtosis-autocorrelation combinations estimated from observed time series. The upper-left panel contains them for 27 daily return series of the most frequently traded stocks in the Stockholm Stock Exchange. These series are also considered in [40]. There seems to be plenty of variation among the series. A large majority have an unreachable combination of κ_4 and ρ_1 in the sense that the combinations do not correspond to a GARCH(1,1) with a finite variance ($\alpha_1 + \beta_1 < 1$). Only four observations appear in the area defined by $\alpha_1 + \beta_1 < 0.999$. The upper-right panel gives a less variable picture. The rates of return are the 20 subperiods of the return series of the S&P 500 index discussed in Section 2. Three of them do not appear in the panel because their kurtosis is too large. All but two of the remaining 17 lie out of reach for the GARCH(1,1) model with normal errors. The lower-left panel tells a similar story. The rates of return are 34 subseries of five major exchange rates, the Japanese yen, the German mark, the English pound, the Canadian dollar, and the Australian dollar, all against the U.S. dollar, from 2 April 1973 to 10 September 2001. One of them, the first subseries of the Canadian dollar, does not appear in the panel because the autocorrelation is 0.456. The lower-right panel contains all data-points in the three other panels. It is seen from the figure that a majority of the points lie even below the lowest isoquant $\alpha_1 + \beta_1 = 0.999$. An obvious conclusion is that the GARCH(1,1) model with normal errors cannot in a satisfactory fashion reproduce the stylized fact of high kurtosis and low-starting autocorrelation of squares observed in a large number of financial series. This is true at least if we require the existence of the unconditional fourth moment of ε_t. We shall return to this point in Section 6.

It is seen from Figure 2 that the first-order autocorrelation of ε_t^2 does decrease with $\alpha_1+\beta_1$ when the kurtosis is kept constant. This may suggest that an integrated GARCH model of [14] could offer an adequate description of the stylized fact. The first-order IGARCH model is obtained by setting $\alpha_1 + \beta_1 = 1$ in (2), which implies that the GARCH process does not have a finite variance. Because there are no moment results to rely on, this possibility was investigated by simulation. Figure 4 contains the same isoquants as before, completed with 100 kurtosis-autocorrelation combinations obtained by simulating the first-order IGARCH with $\beta_1 = 0.9$. The number of observations increases from $T = 100$ in the upper-left panel to 10000 in the lower-right one. It is quite clear that for $T = 100$, it is difficult to even argue that the observations come from a GARCH model. For about one half of the observations the estimated kurtosis lies below three, and for a third, the first-order autocorrelation of squared observations is negative. One can conclude from this that when the null of no conditional heteroskedasticity is rejected for the errors of a macroeconomic equation, estimated using a small number of quarterly observations, fitting an ARCH or a GARCH model to the errors without a close scrutiny of the residuals is hardly a sensible thing to do.

Another conclusion, relevant for our stylized fact considerations, is that when the number of observations increases, the point cloud in the figure moves to the right. This is what it should do since the fourth moment of ε_t does not exist. However, the points follow the isoquants on their way out of the frame, and they do not cross the area where most of the observations in Figure 3 were found. This small simulation experiment thus indicates that the IGARCH model cannot be the solution to the problem that the GARCH(1,1) model with normal errors does not accord with this particular stylized fact.

In applications it is customary not to assume normal errors for z_t in (1) but rather make use of a leptokurtic error distribution such as the t-distribution. Why this is the case can be seen from Figure 5. It contains the same isoquants as before, measured by $\alpha_1 \nu_2 + \beta_1$. This is the condition for covariance stationarity just as $\alpha_1+\beta_1 < 1$ in Figure 1 is in the case of normal errors. It depends on the degrees of freedom of the t-distribution through ν_2. In the left-hand panel the t-distribution has seven degrees of freedom so that $\kappa_4 = 5$ and in the right-hand panel five, in which case $\kappa_4 = 9$. Figure 5 also contains the kurtosis/autocorrelation combinations for the series shown in the fourth panel of Figure 3 but now under the assumption that the errors have a t-distribution with seven (left panel) and five degrees of freedom (right panel). It is seen how the baseline kurtosis now increases from three to five (left panel) and nine (right panel). The observations now fall inside the fan of isoquants, and the corresponding GARCH(1,1) model with the finite fourth moment appears sufficiently flexible to characterize the stylized fact of high kurtosis and low autocorrelation of squared observations.

4.2. EGARCH(1,1) Model

The GARCH(1,1) model with normal errors does not adequately describe the stylized fact of high kurtosis/low autocorrelation of squares combinations. In this section we consider the situation in the symmetric EGARCH(1,1) model. The relationship between κ_4 and ρ_1 for three symmetric EGARCH(1,1) models, $\phi = 0$, with normal errors with different persistence measured by β is depicted in Figure 6[†]. The isoquants now contain the points with β being a constant, while ψ is changing. The kurtosis is a monotonically increasing function of ψ. This figure shows that large values of κ_4 and low values of ρ_1 cannot exist simultaneously for the symmetric EGARCH(1,1) model either. The lowest values for ρ_1 are obtained when β is close to one but these values are not sufficiently low to reach down where the data-points are.

[42] recommended using the Generalized Error Distribution (GED(υ)) for the errors. [22] used the double exponential (Laplace) distribution. The GED(υ) includes both the normal distribution, $\upsilon = 2$, and the Laplace distribution, $\upsilon = 1$, as special cases. If $\upsilon \leq 1$, restrictions on ψ (and ϕ) are needed to guarantee finite moments. Note that the t-distribution for the errors may imply an infinite unconditional variance for $\{\varepsilon_t\}$. For a detailed discussion, see [42]. The autocorrelations of $\{\left|\varepsilon_t\right|^{2m}\}$ with z_t ~GED(υ) can be found in [27].

4.3. ARSV(1) Model

In order to complete our scrutiny of the kurtosis/autocorrelation relationship we consider the first-order ARSV model. [8] have also done similar work. The autocorrelation function of $\{\varepsilon_t^2\}$ of the ARSV(1) model can be expressed as a function of the kurtosis as follows:

$$\rho_n(1) = \frac{(\kappa_4/\kappa_4(z_t))^{\beta^n} - 1}{\kappa_4 - 1}, n \geq 1. \tag{21}$$

Note the similarity between (21) and the corresponding expression for the EGARCH(1,1) model with $\psi = 0$ in footnote 1. In fact, a comparison of these expressions shows that the autocorrelations for this special EGARCH model with normal errors for the same value β are always greater than the corresponding autocorrelations for the ARSV(1) model. Figure 7 contains a plot of the relationship between κ_4 and $\rho_1(1)$ for three ARSV(1) models with normal errors ($\kappa_4(z_t) = 3$) with different persistence measures β. The isoquants now consist of the points with β being 0.95, 0.99, 0.999, respectively, while σ_η^2 is changing. The kurtosis is

---

[†]For the EGARCH(1,1) model with $\psi = 0$ and standard normal errors we can express the autocorrelation function of squared observations as a function of kurtosis: $\rho_n(1) = \frac{(1+\phi^2)(\kappa_4/3)^{\beta^n} - 1}{\kappa_4 - 1}$.

a monotonically increasing function of σ_η^2. An important difference between the symmetric EGARCH(1,1) model and the ARSV(1) model lies in the behaviour of the first-order autocorrelation when the kurtosis is held constant. In the EGARCH(1,1) model, the value of the autocorrelation decreases as a function of β_1, the parameter that controls the decay rate of the autocorrelations. In the ARSV(1) model this value increases as a function of the corresponding parameter β. Thus, contrary to the symmetric EGARCH model, a low first-order autocorrelation and high persistence can coexist in the ARSV model. In general, the first-order autocorrelations, given the kurtosis, are lower in the ARSV than the EGARCH model with normal errors. This may at least partly explain the fact that in some applications the ARSV(1) model seems to fit the data better than its EGARCH or GARCH counterpart. It may also explain the stylized fact mentioned in [48] that β estimated from an ARSV(1) model tends to be lower than the sum $\alpha_1 + \beta_1$ estimated from a GARCH(1,1) model.

In Figure 8 the errors of the ARSV model have a t-distribution with seven (left panel) and five degrees of freedom (right panel). It is seen that when the number of degrees of freedom in the t-distribution decreases, the persistence parameter β only has a negligible effect on the first-order autocorrelation. At the same time, the value of the autocorrelation rapidly decreases with the number of degrees of freedom for any given σ_η^2. Compared to the GARCH(1,1) model, the difference is quite striking.

5. Taylor Effect

5.1. GARCH(1,1) Model

As discussed in Section 2, a large number of financial series display an autocorrelation structure such that the autocorrelation of $\left|\varepsilon_t\right|^{2m}$ decay slowly and the autocorrelations as a function of $m > 0$ peak around $m = 0.5$. [25] defined the corresponding theoretical property and called it the Taylor property. From the results in Section 3 it follows that the existence of the Taylor property in the EGARCH(1,1) and ARSV(1) models can be considered analytically because the analytic expressions for $\mathsf{E}\left|\varepsilon_t\right|^{2m}$ exist for any $m > 0$. This is not true for most GARCH models, however, because analytic expressions are available only for integer moments. An exception is the power-GARCH model of [12]. For this model, certain non-integer moments have an analytic definition, but then, the integer moments generally do not; see [26] and [29].

It is possible to consider a more restricted form of definition that only concerns the first and second moment. The model is then said to have the Taylor property if

$$\rho(\left|\varepsilon_t\right|, \left|\varepsilon_{t-n}\right|) > \rho(\left|\varepsilon_t\right|^2, \left|\varepsilon_{t-n}\right|^2), n \geqslant 1. \tag{22}$$

This choice can be defended by referring to the original discussion in [50]. The problem is that for the standard GARCH model, an analytic definition of $\mathsf{E}\left|\varepsilon_t\right|$ as a function of the parameters is not available. On the other hand, it exists for the AVGARCH(1,1) model defined by [50] and [46]. This prompted [25] to discuss the existence of the Taylor property in the AVGARCH(1,1) model. Their conclusion, based on considerations with $n = 1$ in (22), was that the AVGARCH model possesses the Taylor property if the kurtosis of the model is sufficiently large. However, the difference between the autocorrelations of $\left|\varepsilon_t\right|$ and ε_t^2 remains very small even when the kurtosis is very large. These authors also investigated the existence of the Taylor property in the standard GARCH(1,1) model by simulation, and their results suggested that this model does not have the Taylor property. Of course, due to sample uncertainty, the GARCH model can still generate realizations displaying the Taylor effect, at least when the number of observations is relatively small. This would not, however, happen at the frequency with which the Taylor effect is found in financial series; see [20].

5.2. EGARCH(1,1) Model

We extend the considerations in [25] to the EGARCH(1,1) and ARSV(1) model. For these models, the situation is different. The results of Section 3 allow us to say something about the capability of the EGARCH(1,1) model to generate series with the Taylor property. Figure 9 contains a description of the relationship between κ_4 and the two first-order autocorrelations $\rho_1(m)$, $m = 1, 0.5$, for $\beta = 0.95$ and $\beta = 0.99$. It is seen that the Taylor property is present at high values of the kurtosis. The values of the kurtosis for which the Taylor property is present decrease as a function of β. The difference between the two first-order autocorrelations is substantially greater than in the AVGARCH(1,1) model.

As analytical expressions for non-integer moments of $\mathsf{E}\left|\varepsilon_t\right|^{2m}$, $m > 0$, exist for the EGARCH model, we can extend our considerations by use of them. Figure 10 contains graphs showing the first-order autocorrelation as a function of the exponent m for $\beta = 0.95$ and $\beta = 0.99$ at three different kurtosis values. It turns out that for the symmetric EGARCH process, with kurtosis of the magnitude found in financial time series, the maximum appears to be attained for m around 0.5. The conclusion is that the Taylor property is satisfied for an empirically relevant subset of EGARCH(1,1) models.

5.3. ARSV(1) Model

In order to complete our discussion about Taylor effect we consider the first-order ARSV model. Figure 11 illustrates the relationship between κ_4 and the two first-order autocorrelations $\rho_1(m)$, $m = 1, 0.5$, for $\beta = 0.95$ and $\beta = 0.99$. It is seen that the Taylor property is present already at low values of the kurtosis.

Analogously to the preceding subsection, Figure 12 contains a graph showing the first-order autocorrelation as a function of m for $\beta = 0.95$ and $\beta = 0.99$ and the three different kurtosis values. There is a difference between the EGARCH(1,1) model and the ARSV(1) model regarding the peak value of $\rho_1(m)$ when the persistence parameter changes. In the EGARCH(1,1) model, the peak of the autocorrelation moves to left with higher β_1. In the ARSV(1) model, increasing the value of the corresponding parameter β shifts the peak of the autocorrelation to the right. This feature demonstrates the difference in the relationship between the persistence and the first-order autocorrelation in these two models. Nevertheless, the general conclusion even here is that for the ARSV(1) model, there exists an empirically relevant subset of these models such that the definition of the Taylor property is satisfied.

Thus both the ARSV(1) and the EGARCH(1,1) model appear to reproduce this stylized fact considerably better than the first-order GARCH model.

6. Confidence Regions for the Kurtosis-auto-correlation Combination

When the kurtosis-autocorrelation combination and volatility models were discussed in Section 4, the observations were treated as fixed for simplicity. In reality, they are estimates based on time series. This being the case, it would be useful to account for the uncertainty of these estimates and see whether or not that would change the conclusions offered in Section 4. For this purpose one has to estimate confidence regions for kurtosis-autocorrelation combinations.

The problem then is that it is not possible to obtain these confidence regions analytically. The kurtosis and first-order autocorrelation of squared observations are nonlinear functions of the parameters of the model, be that a GARCH, an EGARCH or an ARSV model. Furthermore, there is no one-to-one mapping between the two parameters of interest and the parameters in the three models. This implies that the confidence regions have to be obtained by simulation. As an example, suppose that the true model generating the time series is a GARCH(1,1) one with a finite fourth moment and fit this model to the series. Use the formulas (5) and (7) to obtain the plug-in estimate of the kurtosis-autocorrelation combinations. Next, use the asymptotic distribution of the maximum likelihood estimator of the parameters and the same formulas to obtain a random sample of kurtosis-autocorrelation combinations from this distribution. The elements that fail the fourth-order moment condition are discarded, and the remaining ones are used for constructing confidence intervals.

In order to illustrate the situation, consider Figure 13 that contains 200 kurtosis-autocorrelation combinations generated from an estimated GARCH(1,1) model. The original time series has been generated from a GARCH(1,1) model with parameters $\alpha_0 = 0.05, \alpha_1 = 0.19121, \beta_1 = 0.75879$ ($\alpha_1 + \beta_1 = 0.95$). A striking feature is that the point cloud has a form of a boomerang that appears to be shaped by the isoquants also included in the figure. This feature has an important consequence: estimating the joint density function of the two variables, kurtosis and autocorrelation estimators, is hardly possible by applying a bivariate kernel estimator based on a linear grid. Such a grid would, however, cover vast areas where no observations are located. Kernel estimation can instead be carried out by replacing the linear grid by a particular nonlinear one that makes use of the isoquants; see [13] for details. Desired confidence intervals are then obtained as highest density regions; [see, for example 53, Section 15.2] and for computational details, [31].

As an application we consider two daily return series of stocks traded in the Stockholm stock exchange. For the stock Assi D with 1769 observations the estimated kurtosis equals 5.8, and the first-order autocorrelation of squared returns equals 0.305. The solid square in Figures 14, 15 and 16 represents this kurtosis/autocorrelation pair. After estimating the three models, the plug-in estimate of the kurtosis/autocorrelation pair can be obtained for each model, and the solid circle represents the estimated pair in the three figures. To estimate the ARSV model we use the quasi-maximum likelihood estimator suggested in [17]. Finally, the solid lines are the 90% confidence regions of the true kurtosis/autocorrelation pair.

For the GARCH(1,1) model in Figure 14 the deviation of the plug-in estimated kurtosis/autocorrelation point from the directly estimated pair is quite small, and the directly estimated combination remains inside the 90% confidence region. Both for the EGARCH model in Figure 15 and for the ARSV model in Figure 16 the plug-in estimate of the first-order autocorrelation is clearly lower than the nonparametric estimate. However, for the EGARCH(1,1) model in Figure 15 the nonparametrically estimated combination remains inside the 90% confidence region, whereas this is not the case for the ARSV(1) model, see Figure 16. The result for the ARSV model is probably due to the fact, discussed in Section 4.3, that the persistence parameter β does not have a prominent role to play in the determination of autocorrelations of squared observations.

Next we consider another return series that has a combination of kurtosis and first-order autocorrelation of squares that lies below even the lowest isoquants for the GARCH model in Figure 3 and the EGARCH model in Figure 6. This is the return series of 2984 observations for the stock SEB that has kurtosis 18.0 and the first-order autocorrelation of squares 0.267. If it is assumed that the series is generated from a GARCH(1,1) model with normal errors, it is seen from Figure 17 that this leads to a low estimate of the kurtosis and the autocorrelation. The kurtosis-autocorrelation combination is heavily underestimated. The 90% confidence region does not cover the nonparametrically estimated kurtosis-autocorrelation combination. We also find a GARCH(1,1) model with t-distributed errors to this series and estimated the 90% confidence region kurtosis-autocorrelation pair under the assumption that the observations are generated by a GARCH(1,1) model. The estimated number of degrees of freedom, $\widehat{\upsilon}$, is close to seven, and the plug-in kurtosis estimate, obtained after rounding $\widehat{\upsilon}$ off to 7, is quite high, equalling 56. It is seen from Figure 18 that the plug-in kurtosis-autocorrelation estimate is not contained in the 90% confidence region. Furthermore, the nonparametric estimate with kurtosis less than 20 and first-order autocorrelation around 0.25 lies far outside the confidence region. It seems that at least in this example, a GARCH(1,1) model with t-distributed errors does not reproduce the stylized facts any better than its counterpart with normal errors.

Similar results are obtained for both the EGARCH model, see Figure 19, and the ARSV model. It is not possible to estimate and graph the corresponding confidence region for the stochastic volatility model because the region turns out to be almost like a section of a one-dimensional curve. It may be noticed, however, that the plug-in kurtosis estimate from the ARSV model is considerably higher than the corresponding estimate from the GARCH model with normal errors, a fact previously emphasized by [8], but lower than the estimate from the GARCH model with t-distributed errors.

A conclusion from this small application, under the assumption that the observations have been generated from a member of the family of models in question, is that the GARCH(1,1) model and the EGARCH(1,1) model cannot reproduce the stylized fact of high kurtosis and low-starting autocorrelation of squares even if we account for the uncertainty. For processes with relatively low kurtosis both the GARCH(1,1) and the EGARCH(1,1) model appear to reproduce the kurtosis-autocorrelation stylized fact better than the first-order ARSV model in the sense that the nonparametrically estimated kurtosis-autocorrelation combination is likely to be covered by the confidence region for the former models but not for the latter one.

7. An Unexplained Stylized Fact

As is clear from the preceding discussion, each one of the three basic models satisfies at least some of the stylized fact considered in this work. There is, however, one frequently encountered feature that cannot be reproduced by any of them: the estimated marginal distribution of many return series is skewed. Such an unconditional distribution cannot be obtained by generalizing the standard GARCH model into an asymmetric one such as the GJR-GARCH [18], QARCH [47] or Smooth Transition GARCH [23, 19, 39] model. For all such models, the unconditional third moment of the process equals zero if it exists as long as the distribution of the error process z_t is symmetric around zero. The same is true for the EGARCH model which has an in-built asymmetric volatility component. If the error distribution of the GARCH model is assumed symmetric, which is a reasonable assumption, a skewed error distribution may be obtained for example by introducing autoregressive structure in the conditional mean; see [30] for discussion. Some researchers have instead assumed that the distribution of the error term z_t is skewed, which also yields conditional skewness. Asymmetry of the error distribution is, however, an unusual assumption, because typically asymmetries in the marginal distribution of the variable to be modelled are parameterized and not “explained”by the error distribution.

It may be argued, however, that the observed asymmetry may often be due to a small number of outliers, for discussion and examples see [34] and [1].

8. Conclusions

In this paper we have shown that there exist possibilities of parameterizing all three models in such a way that they can accommodate and explain many of the stylized facts visible in the data. Even after excluding skewed marginal distributions, some stylized facts may in certain cases remain unexplained. For example, it appears that the standard GARCH(1,1) model may not particularly often generate series that display the Taylor effect. This is due to the fact that this model does not appear to satisfy the corresponding theoretical property, the Taylor property. On the contrary, this property is approximately satisfied for a relevant subset of EGARCH(1,1) and ARSV(1) models and, albeit very narrowly, for a subset of absolute-valued GARCH models.

Many researchers have observed quite early on that for GARCH models, assuming normal errors is too strong a restriction, and they have suggested leptokurtic error distributions in their stead. The results in this paper show how these distributions add to the flexibility of the GARCH model and help the model to reproduce the stylized fact of high kurtosis and relative low autocorrelations of squared observations. It is also demonstrated that the IGARCH model with normal errors does not rescue the normality assumption. As a drawback it may be noted that the parameterization of the first-order autoregressive stochastic volatility model becomes very restrictive when the amount of the leptokurtosis in the error distribution increases, and the model therefore cannot accommodate ’easy’ situations with relatively low kurtosis and high autocorrelations of squared observations.

The paper contains an application of a novel method of obtaining confidence regions for the kurtosis-autocorrelation combinations. The brief application of this method to stock returns indicates, not surprisingly, that when normality of errors is assumed, the GARCH model as well as the EGARCH model are at their best when it comes to characterizing models based on time series with relatively low kurtosis and high first-order autocorrelation of squares. Time series displaying a combination of high kurtosis and high autocorrelation are better modelled using an ARSV(1) model. While this observation may serve as a rough guide when one wants to select one of these models, nonnested tests are also available for comparing them. Examples of such tests have already been mentioned in the Introduction.

Another observation that emerges from the empirical example is that the estimated kurtosis-autocorrelation combination is often an underestimate compared to the one estimated nonparametrically from the data. This is the case when the kurtosis is high and the errors are normal. This fact may be interpreted as support to the notion that a leptokurtic error distribution is a necessity when using GARCH models. This idea is contradicted, however, by the fact that assuming a t-distribution for the errors may at least in some cases lead to a large discrepancy in the opposite direction between the plug-in estimate of the kurtosis and the nonparametric estimate. These results may suggest that daily return series in fact contain truly exceptional observations in the sense that they cannot be satisfactorily explained by the members of the standard GARCH or EGARCH family of models.

This argument receives a certain amount of support from [34] who investigated robust estimation of skewness and kurtosis of return series. It turned out that robust estimates were much less extreme than the standard ones, and removing a small number of outliers from the series considerably lowered the standard kurtosis estimates. Considering the kurtosis-autocorrelation combinations using robust measures of kurtosis and autocorrelation may therefore be a useful addition to the analysis of stylized facts. [54] have recently carried out work in this direction.

The present investigation is only concerned with first-order models, and a legitimate question is whether adding more lags would enhance the flexibility of the models. Such additions would certainly help to generate and reproduce more elaborate autocorrelation patterns for the squared observations than is the case with first-order models. It is far from certain, however, that they would also improve reproduction of the stylized facts considered in this study.

ACKNOWLEDGEMENTS This research has been supported by the Swedish Council for Research in the Humanities and Social Sciences and the Danish National Research Foundation. A considerable part of the work was done while both authors were with the Department of Economic Statistics, Stockholm School of Economics.

References

[1] Cristina Amado and Timo Teräsvirta. Modelling conditional and unconditional heteroskedasticity with smoothly time-varying structure. SSE/EFI Working Paper Series in Economics and Finance 691, Stockholm School of Economics, 2008.

[2] J Andersson. On the normal inverse Gaussian stochastic volatility model. *Journal of Business Economic Statistics*, 19:44–54, 2001.

[3] X Bai, J R Russell, and G C Tiao. Kurtosis of GARCH and stochastic volatility models with non-normal innovations. *Journal of Econometrics*, 114:349–360, 2003.

[4] R T Baillie, T Bollerslev, and H-O Mikkelsen. Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 74:3–30, 1996.

[5] Richard T. Baillie and Claudio Morana. Modeling long memory and structural breaks in conditional variances: An adaptive FIGARCH approach. *Journal of Economic Dynamics and Control*, 33:1577–1592, 2009.

[6] T Bollerslev. On the correlation structure for the generalized autoregressive conditional heteroskedastic process. *Journal of Time Series Analysis*, 9:121–131, 1988.

[7] Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31:307–327, 1986.

[8] M A Carnero, D Peña, and E Ruiz. Persistence and kurtosis in GARCH and stochastic volatility models. *Journal of Financial Econometrics*, 2:319–342, 2004.

[9] Y-T Chen and C-M Kuan. The pseudo-true score encompassing test for non-nested hypotheses. *Journal of Econometrics*, 106:271–295, 2002.

[10] C S-J Chu. Detecting parameter shift in GARCH models. *Econometric Reviews*, 14:241–266, 1995.

[11] Rainer Dahlhaus and Suhasini Subba Rao. Statistical inference for time-varying ARCH processes. *Annals of Statistics*, 34:1075–1114, 2006.

[12] Z Ding, C W J Granger, and R F Engle. A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, 1:83–106, 1993.

[13] Bruno Eklund. Estimating confidence regions over bounded domains. *Computational Statistics and Data Analysis*, 49:349–360, 2005.

[14] R F Engle and T Bollerslev. Modelling the persistence of conditional variances. *Econometric Reviews*, 5:1–50, 1986.

[15] R F Engle and V K Ng. Measuring and testing the impact of news on volatility. *Journal of Finance*, 48:1749–1778, 1993.

[16] Robert F. Engle and José Gonzalo Rangel. The spline GARCH model for low-frequency volatility and its global macroeconomic causes. *Review of Financial Studies*, 21:1187–1222, 2008.

[17] E Ghysels, A C Harvey, and E Renault. Stochastic volatility. In G S Maddala and C R Rao, editors, *Handbook of Statistics*, volume 14, pages 119–191, Amsterdam, 1996. North Holland.

[18] L Glosten, R Jagannathan, and D Runkle. On the relation between expected value and the volatility of the nominal excess return on stocks. *Journal of Finance*, 48:1779–1801, 1993.

[19] G González-Rivera. Smooth transition GARCH models. *Studies in Non-linear Dynamics and Econometrics*, 3:61–78, 1998.

[20] C W J Granger and Z Ding. Some properties of absolute returns: An alternative measure of risk. *Annales d'économie et de statistique*, 40:67–95, 1995.

[21] C W J Granger, M L King, and H White. Comments on testing economic theories and the use of model selection criteria. *Journal of Econometrics*, 67:173–187, 1995.

[22] C W J Granger, S Spear, and Z Ding. Stylized facts on the temporal and distributional properties of absolute returns: An update. In Chan W-S, W K Li, and H Tong, editors, *Statistics and Finance: An Interface*, pages 97–120, London, 2000. Imperial College Press.

[23] G E Hagerud. *A New Non-Linear GARCH Model*. Stockholm: EFI Economic Research Institute, 1997.

[24] C He and T Teräsvirta. Fourth moment structure of the GARCH(p,q) process. *Econometric Theory*, 15:824–846, 1999.

[25] C He and T Teräsvirta. Properties of moments of a family of GARCH processes. *Journal of Econometrics*, 92:173–192, 1999.

[26] C He and T Teräsvirta. Statistical properties of the asymmetric power ARCH process. In R F Engle and H White, editors, *Cointegration, Causality, and Forecasting. Festschrift in Honour of Clive W J Granger*, pages 462–474, Oxford, 1999. Oxford University Press.

[27] C He, T Teräsvirta, and H Malmsten. Moment structure of a family of first-order exponential GARCH models. *Econometric Theory*, 18:868–885, 2002.

[28] Changli He. Moments and the autocorrelation structure of the exponential GARCH(p,q) process. SSE/EFI Working Paper Series in Economics and Finance 359, Stockholm School of Economics, 2000.

[29] Changli He, Hans Malmsten, and Timo Teräsvirta. Higher-order dependence in the general power ARCH process and the role of the power parameter. In Shalabh and C. Heumann, editors, *Recent Advances in Linear Models and Related Areas*, pages 231–251, New York, 2008. Springer.

[30] Changli He, Annastiina Silvennoinen, and Timo Teräsvirta. Parameterizing unconditional skewness in models for financial time series. *Journal of Financial Econometrics*, 6:208–230, 2008.

[31] R Hyndman. Computing and graphing highest density regions. *American Statistician*, 50:120–126, 1996.

[32] M Karanasos and J Kim. Moments of the ARMA-EGARCH model. *Econometrics Journal*, 6:146–166, 2003.

[33] S Kim, N Shephard, and S Chib. Stochastic volatility: Likelihood inference and comparison with ARCH models. *Review of Economic Studies*, 65:361–393, 1998.

[34] T-H Kim and H White. On more robust estimation of skewness and kurtosis. *Finance Research Letters*, 1:56–73, 2004.

[35] J-H Lee and B W Brorsen. A non-nested test of GARCH vs. EGARCH models. *Applied Economics Letters*, 4:765–768, 1997.

[36] R Liesenfeld and R C Jung. Stochastic volatility models: Conditional normality versus heavy-tailed distributions. *Journal of Applied Econometrics*, 15:137–160, 2000.

[37] S-J Lin and J Yang. Testing shift in financial models with conditional heteroskedasticity: An empirical distribution function approach. 1999. Research Paper 30, University of Technology Sydney, Quantitative Finance Research Group.

[38] S Ling and M McAleer. Testing GARCH versus E-GARCH. In Chan W-S, W K Li, and H Tong, editors, *Statistics and Finance: An Interface*, pages 226–242, London, 2000. Imperial College Press.

[39] S Lundbergh and T Teräsvirta. Evaluating GARCH models. *Journal of Econometrics*, 110:417–435, 2002.

[40] Hans Malmsten. Evaluating exponential GARCH models. SSE/EFI Working Paper Series in Economics and Finance 564, Stockholm School of Economics, 2004.

[41] T Mikosch and C Stărică. Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects. *Review of Economics and Statistics*, 86:378–390, 2004.

[42] D B Nelson. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59:347–370, 1991.

[43] D B Nelson and C Q Cao. Inequality constraints in the univariate GARCH model. *Journal of Business and Economic Statistics*, 10:229–235, 1992.

[44] S H Poon and C W J Granger. Forecasting volatility in financial markets. *Journal of Economic Literature*, 41:478–539, 2003.

[45] T Rydén, T Teräsvirta, and S Åsbrink. Stylized facts of daily return series and the hidden Markov model. *Journal of Applied Econometrics*, 13:217–244, 1998.

[46] G W Schwert. Why does stock market volatility change over time? *Journal of Finance*, 45:1129–1155, 1989.

[47] E Sentana. Quadratic ARCH models. *Review of Economic Studies*, 62:639–661, 1995.

[48] N Shephard. Statistical aspects of ARCH and stochastic volatility. In D R Cox, O E Bandorff-Nielsen, and D V Hinkley, editors, *Statistical Models in Econometrics, Finance, and Other Fields*, pages 1–67, London, 1996. Chapman and Hall.

[49] Neil Shephard and Torben G. Andersen. Stochastic volatility: Origins and overview. In Torben G. Andersen, Richard A. Davis, Jens-Peter Kreiss, and Thomas Mikosch, editors, *Handbook of Financial Time Series*, pages 233–254, New York, 2009. Springer.

[50] S Taylor. *Modelling Financial Time Series*. Wiley, New York, 1986.

[51] T Teräsvirta. Two stylized facts and the GARCH(1,1) model. Working Paper Series in Economics and Finance, Stockholm School of Economics, No. 96, 1996.

[52] Timo Teräsvirta. An introduction to univariate GARCH models. In Torben G. Andersen, Richard A. Davis, Jens-Peter Kreiss, and Thomas Mikosch, editors, *Handbook of Financial Time Series*, pages 17–42, New York, 2009. Springer.

[53] Timo Teräsvirta, Dag Tjøstheim, and Clive W. J. Granger. *Modelling Nonlinear Economic Time Series*. Oxford University Press, Oxford, 2010.

[54] Timo Teräsvirta and Zhenfang Zhao. Stylized facts of return series, robust estimates, and three popular models of volatility. SSE/EFI Working Paper Series in Economics and Finance 662, Stockholm School of Economics, 2007.

[55] Y K Tse, X B Zhang, and J Yu. Estimation of hyperbolic diffusions using Markov Chain Monte Carlo method. *Quantitative Finance*, 4:158–169, 2004.

[56] Sébastien van Bellegem and Rainer von Sachs. Forecasting economic time series with unconditional time-varying variance. *International Journal of Forecasting*, 20:611–627, 2004.

Appendix

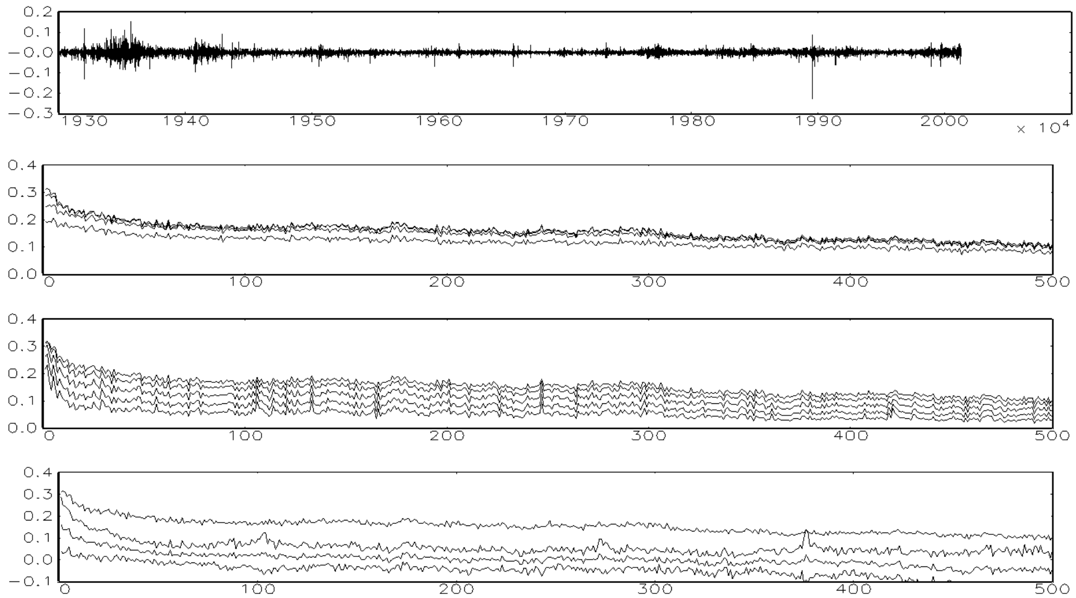

Figure 1: Uppermost panel, log-returns of the S&P 500 index 3 January 1928 to 19 September 2001. Second panel, the autocorrelation function of $|r_t|^m$, $m = 0.25, 0.5, 0.75, 1$, from low to high, for the S&P 500 index. Third panel, the autocorrelation function of $|r_t|^m$, $m = 1, 1.25, 1.5, 1.75, 2$, from high to low, for the S&P 500 index. Lowest panel, the autocorrelation function of $|r_t|$ for the whole series (highest graph) and the mean of the corresponding autocorrelations of the 20 equally long subseries of the S&P 500 index together with the plus/minus one standard deviation band.

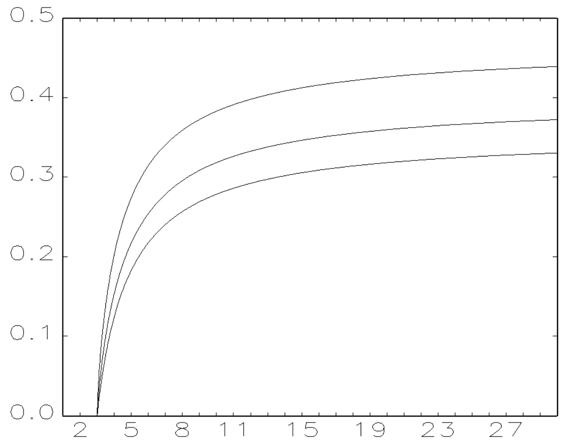

Figure 2: Combinations of the first-order autocorrelation of squared observations and kurtosis for the GARCH(1,1) model with normal errors for various values of $\alpha+\beta$. Isoquants from lowest to highest: $\alpha+\beta=0.999$, 0.99 and 0.95.

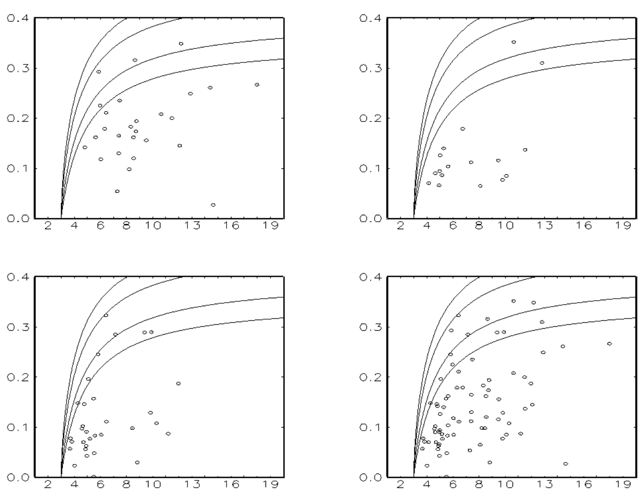

Figure 3: Combinations of the first-order autocorrelation of squared observations and kurtosis for the GARCH(1,1) model with normal errors for various values of $\alpha+\beta$ together with observed combinations of daily rates of return: Upper-left panel, daily returns of the 27 most traded stocks at the Stockholm Stock Exchange. Lower-left panel, the S&P 500 index 3 January 1928 to 19 September 2001, divided to 20 equally long subseries. Upper-right panel, five major daily exchange rates series divided to 34 subseries. Lower-right panel, all observations. Isoquants from lowest to highest: $\alpha+\beta=0.999$, 0.99, 0.95 and 0.9.

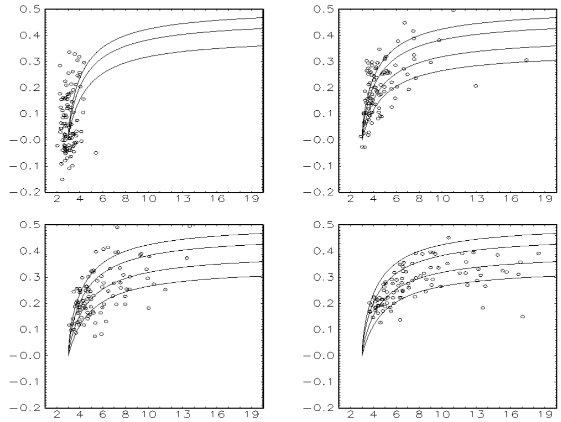

Figure 4: Combinations of the first-order autocorrelation of squared observations and kurtosis for the GARCH(1,1) model with normal errors for various values of $\alpha+\beta$ together with 100 realizations based on T simulated observations from an IGARCH(1,1) model with $\alpha_0 = \alpha = 0.1$: $T = 100$ (first panel), 500 (second panel), 1000 (third panel) and 2000 (fourth panel). Isoquants from lowest to highest: $\alpha + \beta = 0.999$, 0.99, 0.95 and 0.9.

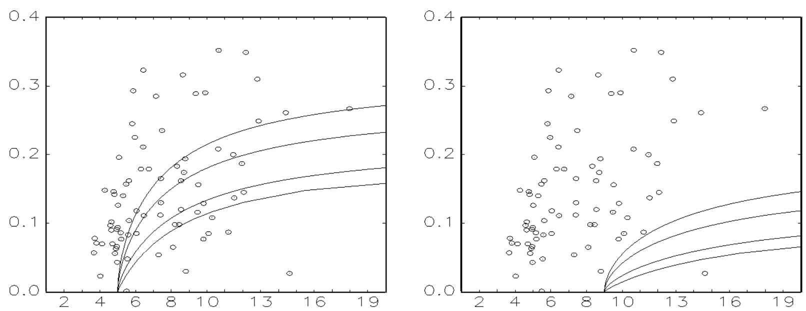

Figure 5: Combinations of the first-order autocorrelation of squared observations and kurtosis for the GARCH(1,1) model with t-distributed errors for various values of $\alpha\nu_2+\beta$: left panel: t(7), right panel: t(5). Isoquants from lowest to highest: $\alpha+\beta=0.999$, 0.99, 0.95 and 0.9. The observed combinations are the same as in the lower-right panel of Figure 1.

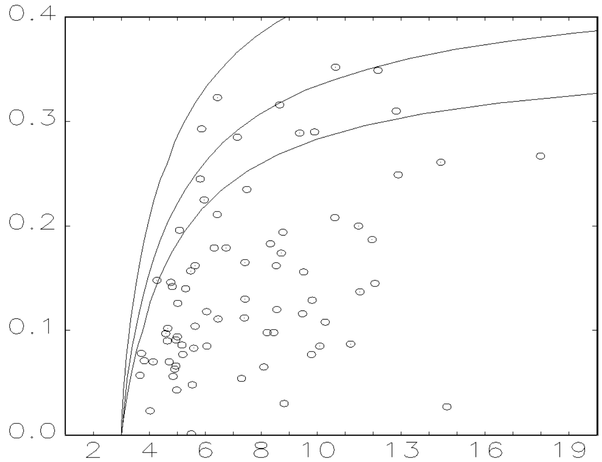

Figure 6: Combinations of the first-order autocorrelation of squared observations and kurtosis for the EGARCH(1,1) model with normal errors for various values of β. The isoquants from lowest to highest: $\beta = 0.95$, 0.99 and 0.999. The observed combinations are the same as in the fourth panel of Figure 1.

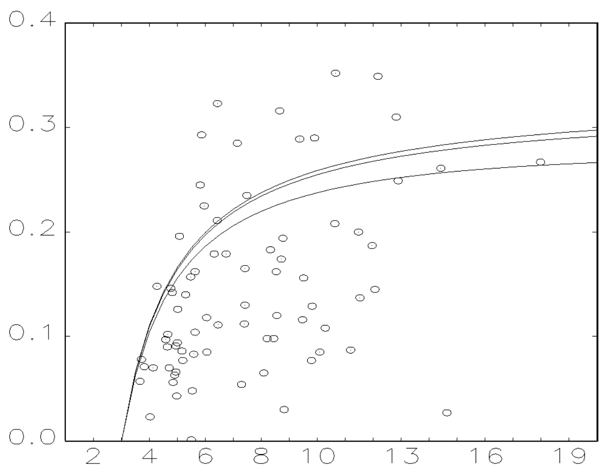

Figure 7: Combinations of the first-order autocorrelation of squared observations and kurtosis for the ARSV(1) model with normal errors for various values of β. Isoquants from lowest to highest: $\beta = 0.999$, 0.99 and 0.95. The observed combinations are the same as in the lower-right panel of Figure 1.

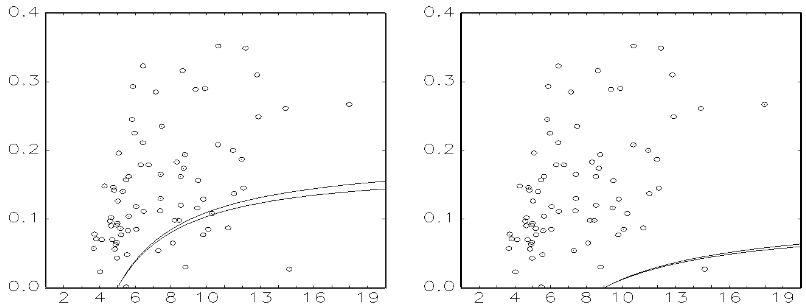

Figure 8: Combinations of the first-order autocorrelation of squared observations and kurtosis for the ARSV(1) model with t-distributed errors for various values of β: left panel: t(7), right panel: t(5). Isoquants from lowest to highest: $\alpha+\beta = 0.99$ and 0.95. The observed combinations are the same as in the lower-right panel of Figure 1.

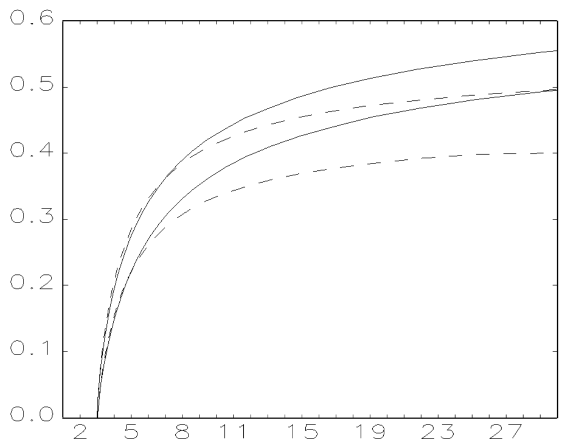

Figure 9: Combinations of two first-order autocorrelations, the squared observations (dashed line) and the absolute observations (solid line), and corresponding kurtosis values for the EGARCH(1,1) model with normal errors for $\beta = 0.99$ (low) and $\beta = 0.95$ (high).

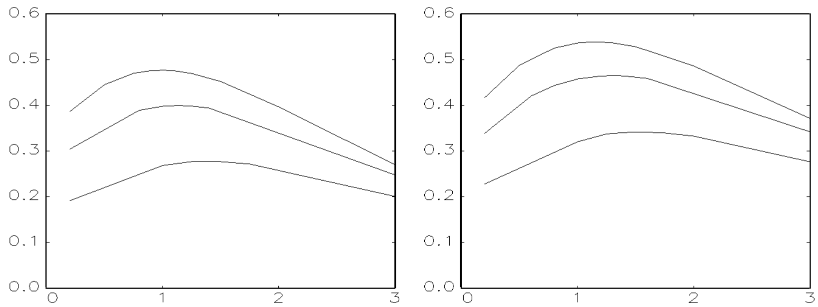

Figure 10: Combinations of the first-order autocorrelation of absolute-valued observations raised to power m as a function of m for the EGARCH(1,1) model with normal errors for $\beta = 0.95$ (left panel) and $\beta = 0.99$ (right panel) at three kurtosis values. From low to high: $\kappa_4 = 6$, 12 and 24.

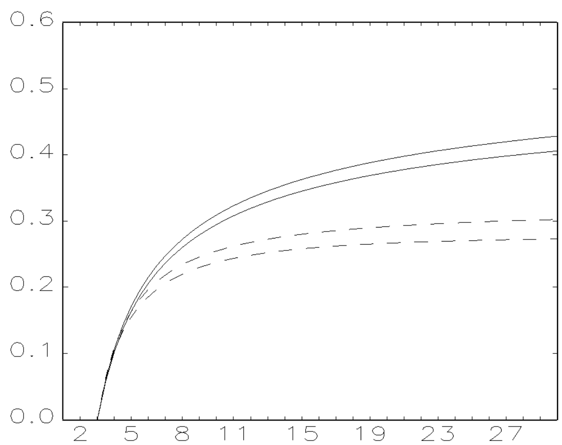

Figure 11: Combinations of two first-order autocorrelations, the squared observations (dashed line) and the absolute observations (solid line), and corresponding kurtosis values for the ARSV(1) model with normal errors for $\beta = 0.95$ (low) and $\beta = 0.99$ (high).

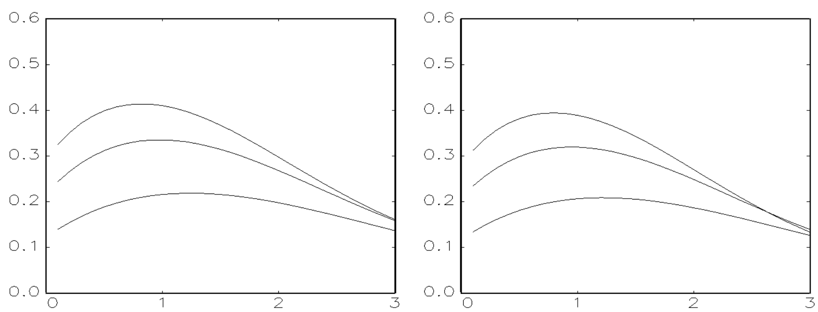

Figure 12: Combinations of the first-order autocorrelation of absolute-valued observations raised to power m as a function of m for the ARSV(1) model with normal errors for $\beta = 0.95$ (left panel) and $\beta = 0.99$ (right panel) at three kurtosis values. From low to high: $\kappa_4 = 6$, 12 and 24.

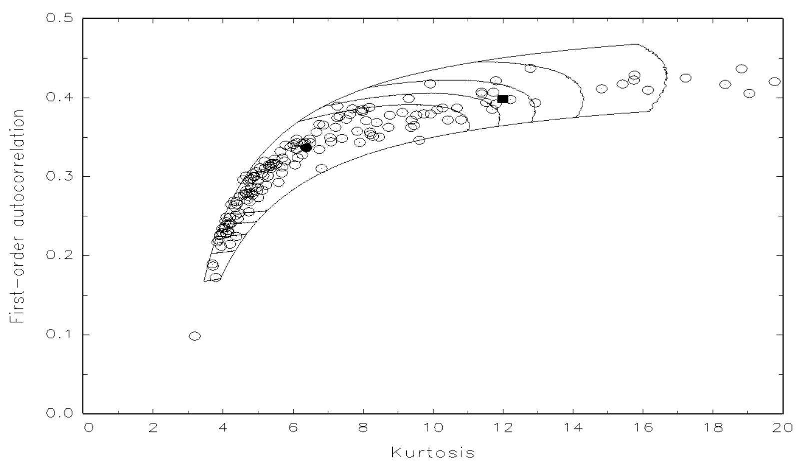

Figure 13: Simulated kurtosis/autocorrelation combinations for the GARCH(1,1) model with $(\alpha_0,\alpha_1,\beta)$=(0.05,0.19121,0.75879), and approximative 50%, 60%, 70%, 80%, and 90% confidence intervals of the true value, 1000 observations and 200 realizations. Solid square is the true value: solid circle is the plug-in estimate; empty circles are generated combinations.

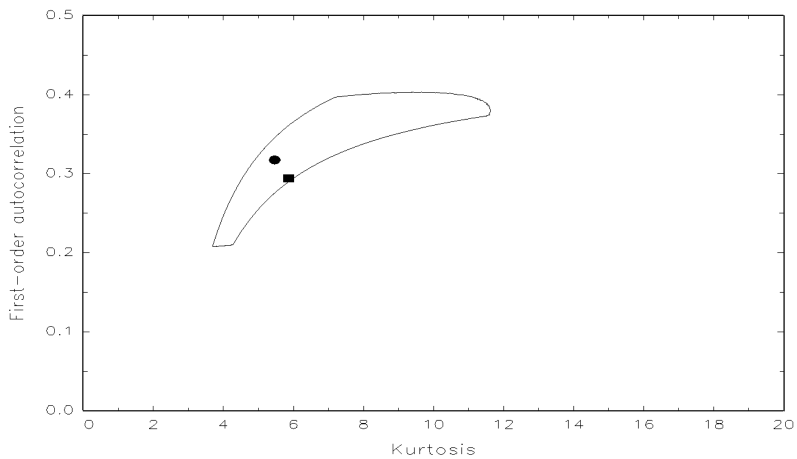

Figure 14: Approximate 90% confidence region based on 200 realizations of the true kurtosis/autocorrelation combination for the Assi D return series under the assumption that the observations have been generated by a GARCH(1,1) model. Solid square is the nonparametrically estimated value, solid circle is the plug-in estimate.

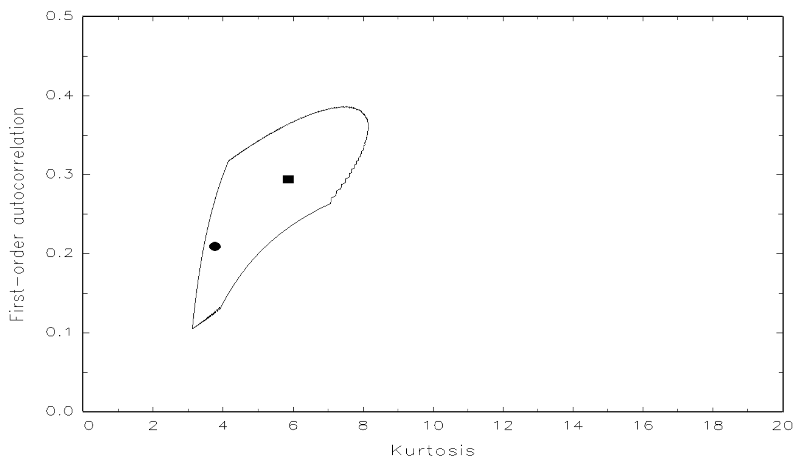

Figure 15: Approximate 90% confidence region based on 200 realizations of the true kurtosis/autocorrelation combination for the Assi D return series under the assumption that the observations have been generated by an EGARCH(1,1) model. Solid square is the nonparametrically estimated value, solid circle is the plug-in estimate.

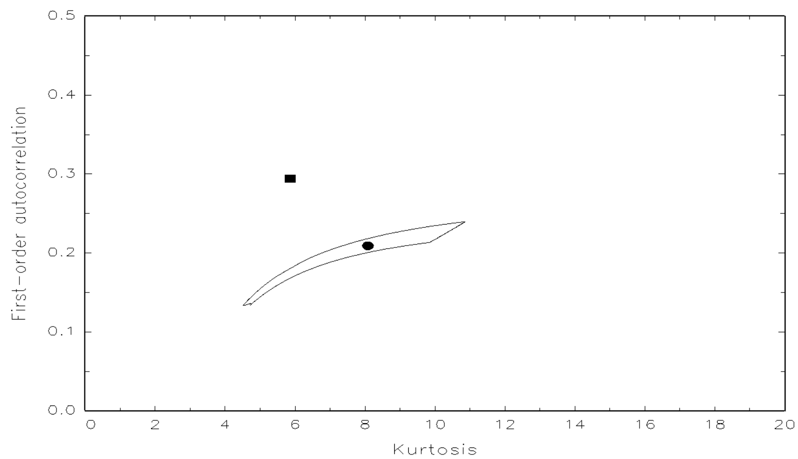

Figure 16: Approximate 90% confidence region based on 200 realizations of the true kurtosis/autocorrelation combination for the Assi D return series under the assumption that the observations have been generated by an ARSV(1,1) model. Solid square is the nonparametrically estimated value, solid circle is the plug-in estimate.

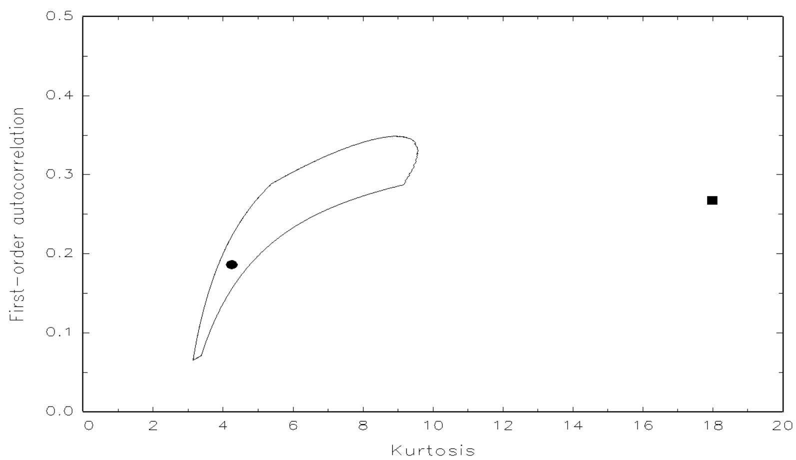

Figure 17: Approximate 90% confidence region based on 200 realizations of the true kurtosis/autocorrelation combination for the SEB return series under the assumption that the observations have been generated by a GARCH(1,1) model. Solid square is the nonparametrically estimated value, solid circle is the plug-in estimate.

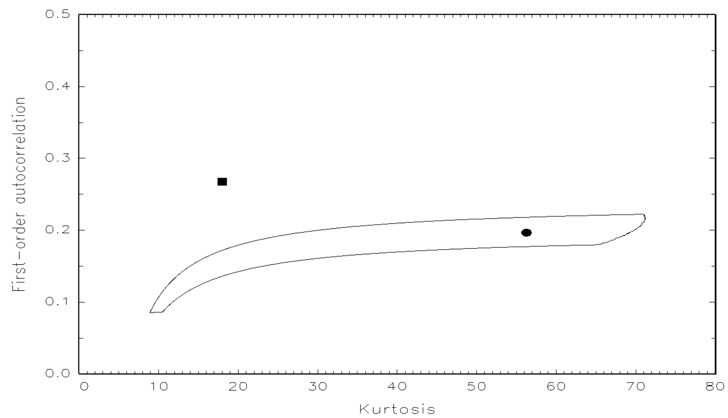

Figure 18: Approximate 90% confidence region based on 200 realizations of the true kurtosis/autocorrelation combination for the SEB return series under the assumption that the observations have been generated by a GARCH(1,1) model with t-distributed errors ($\nu = 7$). Solid square is the nonparametrically estimated value, solid circle is the plug-in estimate.

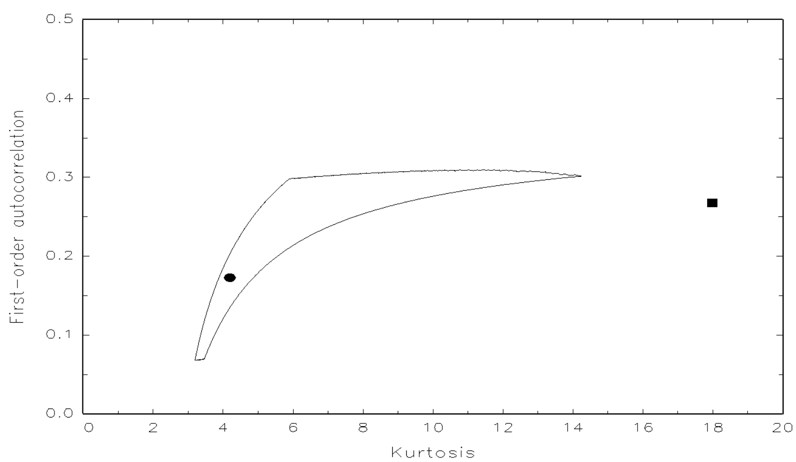

Figure 19: Approximate 90% confidence region based on 200 realizations of the true kurtosis/autocorrelation combination for the SEB return series under the assumption that the observations have been generated by a GARCH(1,1) model with t-distributed errors ($\nu = 7$). Solid square is the nonparametrically estimated value, solid circle is the plug-in estimate.