

Working Papers in Statistics
No 2015:9

Department of Statistics
School of Economics and Management
Lund University

Tail behavior and dependence structure in the APARCH model

FARRUKH JAVED, LUND UNIVERSITY

KRZYSZTOF PODGÓRSKI, LUND UNIVERSITY



TAIL BEHAVIOR AND DEPENDENCE STRUCTURE IN THE APARCH MODEL

Farrukh Javed and Krzysztof Podgórski
SCHOOL OF ECONOMICS AND MANAGEMENT
LUND UNIVERSITY, SWEDEN

September 27, 2015

Abstract

The APARCH model is a generalization of the GARCH model that attempts to capture asymmetric responses of returns and of volatility to positive and negative ‘news shocks’ – the phenomenon known as the leverage effect. Despite its potential, the model’s mathematical properties have not yet been fully investigated. While the capacity of the model to account for the leverage effect is clear from its defining structure, little is known how the effect is quantified in terms of the model’s parameters. The same applies to the quantification of heavy tails and time dependence. Here, in an attempt to fill this void, we study the model in further detail. We obtain sufficient conditions of its existence in different metrics as well as explicit forms of important characteristics: skewness, kurtosis, correlations and leverage. Utilizing these results, we analyze the roles of the parameters and discuss statistical inference. We also propose a natural extension by introducing an additional parameter and discuss how it affects the model. Through theoretical results and a Monte Carlo study we demonstrate that the model can produce heavy-tailed data. We illustrate these properties using S&P500 data as well as country indices for dominant European economies.

1 Introduction

In the field of finance, it has been long observed and exhaustively documented that the data exhibit distinct features that call for more general models than the linear ones based on the Gaussian distribution. Among the most frequently quoted non-Gaussian and nonlinear features are: heavy-tailed distributions, clustering and asymmetries of

The authors were supported by the Riksbankens Jubileumsfond Grant Dnr: P13-1024:1 and the Swedish Research Council Grant Dnr: 2013-5180.

volatility, and to a lesser degree asymmetry in the return distribution (see [7] and references therein).

Heavy-tailedness goes back to Mandelbrot, see [15], who noticed that the price changes of cotton futures showed a much heavier tail than normal. The non-Gaussianity in the data is of serious concern because heavy-tailed extreme values have serious implication for risk management and assessment. It was argued in the literature that misspecified tail behavior in a distributional model can be disastrous for a financial analyst since the price and hedge strategies will not take the large returns into account.

Equally important aspects of the data are the long-range dependence and volatility clustering observed in market returns. To quote from Mandelbrot in [15] "... large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes." This stylized fact is sometimes associated by various authors with empirical evidence that the absolute or squared returns are substantially more correlated than are the returns themselves (see [5] and [19] as well as the references therein). Additionally it has also been observed that the absolute and squared returns display slowly decaying autocorrelation as the lag increases.

It is now commonly accepted that to account for the special properties of financial and, particularly, stock return data, one has to give up either on the linearity of the models, or on the Gaussian distribution of the noise driving a model in question, or both. Over the years, an enormous variety of stochastic models, both discrete and continuous time, have been proposed. The two classical directions are either to treat volatility as non-random conditionally on the past (conditionally heteroscedastic volatility) or to add independent randomness to the volatility model (stochastic volatility); see [19] for a classical overview of general principles and classes of such models. It is worth mentioning that a non-Gaussian noise model can also be viewed as a stochastic volatility model if the noise is a variance normal mixture; see for example [1] and [13]. We do not explore stochastic volatility in this work, instead focusing on the conditionally heteroscedastic volatility model.

One of the first volatility models for financial data was proposed in Engle's seminal paper [6] in an attempt at modeling non-constant volatility effects (volatility clustering) which seemed to be present in the data. The main idea was to model current, unobserved and non-constant variances of returns (volatility) through a form of autoregressive equation in which they are also dependent on the past noise. The same noise is used both in the volatility and to drive the observed returns that follow an autoregressive model conditionally on the unobserved volatility.

It is well known to financial practitioners that the vast majority of data show various systematic asymmetries (see, for example [17] and [2]). Among them two have been subject of more thorough studies, namely asymmetry in the distribution of returns and asymmetry in the way volatility responds to positive and negative (relatively to the mean) returns. For example, asymmetries in empirical distributions of five stock indices and of six foreign exchange rates were observed in [14] and accounted for by a variable skewness and kurtosis time series model. In [20], it has been shown that intro-

ducing asymmetric distribution for the noise allows for more adequate representation of outlying observations.

Another form of asymmetry stems from the fact that the market is prone to react differently to positive as opposed to negative returns. In the financial terminology, the leverage on a company valuation results in an increase in the volatility of the stock price, i.e. the larger leverage, the larger increase in the volatility. It was heavily argued in the literature, (see [3] for some of the earliest work on this topic), that a decrease in stock valuation increases the leverage, as compared to when there is an increase of the stock value. In layman terms, this phenomenon, referred to as the *leverage effect*, means that good and bad news have different predictability for the future volatility, i.e. the effect of positive response on volatility is different than of a negative response of the same size. Such effect can be detected in the data by using the so-called autocoskewness, i.e. the negative and significant correlation between returns and future squared returns, as discussed in [8].

Ding-Granger-Engel's APARCH model

By introducing non-constant, conditionally heteroscedastic volatility, a relatively simple form of non-linearity became available in time series modeling. However, in order to obtain asymmetry due the leverage effect, which exhibits in the ‘bad’ news having a more prominent effect on volatility than the ‘good’ news, some structural changes were needed to the model. In [5], the authors address these issues by modifying the existing model in a way to account for such discrepancies. They make use of the power term in the volatility equation and introduce asymmetry weights for the positive and negative error terms. Such asymmetric power stochastic volatility models have been extensively discussed in the literature and, in principle, have capacity to model important aspects of the data, see, for example, [9].

The model that is referred to as APARCH($\alpha_0, \alpha, \beta, \delta, \theta$), $\alpha_0 > 0$, $\alpha \geq 0$, $\beta \in [0, 1)$, $\theta \in [-1, 1]$, $\delta > 0$, is a generalized version of Bollerslev's GARCH introduced in [4] and can be described through

$$\begin{aligned} y_t &= f(y_{t-1}, y_{t-2}, \dots) + \epsilon_t, & \epsilon_t &= \rho_t e_t, \\ \rho_t^\delta &= \alpha_0 + \rho_{t-1}^\delta \lambda_{t-1}, \end{aligned} \tag{1}$$

where $\lambda_t = \alpha [(1 - \theta)^\delta e_t^{+\delta} + (1 + \theta)^\delta e_t^{-\delta}] + \beta$, variables ρ_t and e_t are independent, with e_t having a standard normal distribution. Here e_t^+ and e_t^- stand for the positive and negative part of e_t , respectively, while the parameters and their role in the model are discussed in the introduction to Section 5. The generic autoregressive function $f(y_{t-1}, y_{t-2}, \dots)$ of the past values is irrelevant for this paper, but one can consider linear autoregressive models of any order, for example $f(y_{t-1}) = a_0 + a_1 y_{t-1}$ leads to AR(1) model. It should be mentioned that the efficient market hypothesis can be realized by taking f as a constant shift μ , as for example in [6], from where the ARCH model originated.

A direct statistical fitting is based on two facts associated with the construction of the model. Firstly, the conditional variance ρ_t^δ to depend on the past realizations represented by $\{y_s, s < t\}$ or, equivalently, by $\{\epsilon_s, s < t\}$. Secondly, while the process ϵ_t is not normally distributed, the process y_t defined in (1) has, conditionally on the past, normal distribution, because ρ_t depends only on the past values $\{y_s, s < t\}$. Therefore, in [5], the joint density is defined by considering all the conditional densities. As a result, the parameters of the model can be estimated by maximizing the log-likelihood as the sum of the conditional log-likelihoods corresponding to (1) by retrieving the ‘estimated’ volatility values through the recursive relation. To proper statistical terminology we refer to the values of volatility obtained in this manner as the empirical volatility.

The conditional Gaussian likelihood works best under the assumption that e_t are indeed normally distributed. However in the real data, as exemplified below, the standardized residuals \hat{e}_t based on this model and computed from the data continue to show some non-Gaussian features.

A motivating example

Here we aim at analyzing the properties of residuals from the model based on the same S&P500 data as used in the original work [5]. We complement the analysis by using simulated data from the fitted model and under the gaussian paradigm. We are mainly interested in the tail behavior of the residuals. The latter are obtained by using the maximum likelihood method with the Gaussian likelihood conditionally on the volatility as described above. The simulation is done using the parameter estimates presented in [5]. Below we report the kurtosis of residuals ($\hat{\epsilon}_t$) and standardized residuals, i.e. residuals scaled by empirical volatility ($\hat{e}_t = \hat{\epsilon}_t/\hat{\rho}_t$).

Table 1: Analysis of heavy tailed behavior via kurtosis.

DATA	KURTOSIS	
S&P500 data	Returns	26.12
	Residuals	24.89
	Stnd. Residuals	8.18
Simulation	Returns	7.33
	Residuals	7.23
	Stnd. Residuals	2.96

If in the model (1) we assume a Gaussian noise, the non-linear component ρ_t^δ can help to capture some heavy-tailedness in the data. However, as our analysis of S&P 500 data shows, it may be not sufficient to address the actual heavy-tailedness of residuals.

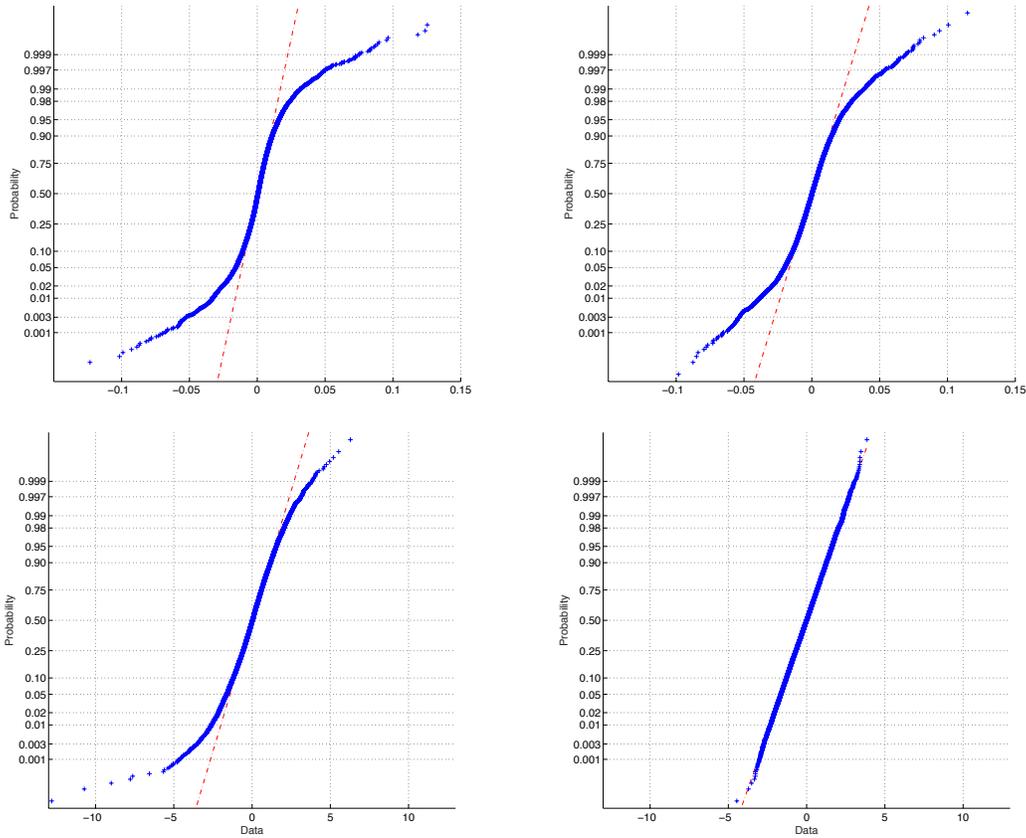


Figure 1: Normal probability plots: (*Top-left*) the SP data; (*Top-right*) simulated data from the model with fitted parameters; (*Bottom-left*) the residuals from the real data after being standardized by the empirical volatility; (*Bottom-right*) the standardized residuals from the simulated data.

First consider the case of fitting the model to our data. From the tabulated values of kurtosis reported in Table 1, it can be seen that the kurtosis for real returns' data and the estimated residuals are, as expected, close to each other and show very large kurtosis (26.12 and 24.89 respectively). This serves as an evidence of a quite heavy tailed unconditional distribution. The key observation, however, is that even though the model removed significant amount of mass from the tail, the estimated standardized residuals still show excess kurtosis with a value of the kurtosis coefficient equals to 8.18.

Next, we apply the same procedure to simulated data that are generated from the model with the set of parameters taken from a fit to the real data. The value of kurtosis coefficient for simulated returns is 7.33. This value, while showing clear departure from the normal distribution, is far below the kurtosis of real returns' data (26.12), which indicates that the simulated data couldn't even generate the magnitude of tails as seen in the real data. However, the kurtosis for the estimated standardized residuals is very

close to 3, which is expected since the simulation was actually done from the Gaussian distribution. We illustrate it graphically by presenting the qq-plots of the real and simulated data together as well as of the estimated residuals and the standardized residuals. The graphs are presented in Figure 1. These observations exemplify that the model is incapable of fully accounting for heavy-tailedness in the data. To remedy this, in this work, we promote an approach that makes a fit that preserves the sample kurtosis. Possibly, a non-Gaussian distribution of the standardized innovations can produce a more accurate models for this data set, however, this will not be pursued in this work, see also [13] and [12].

Contributions

Our example reveals that the degree to which various components of the APARCH model encompass the so-called ‘stylized facts’ (heavy tails, asymmetry, volatility clustering, long-range dependence etc.), is not evident and thus has to be thoroughly investigated. Surprisingly, there is a lack of comprehensive accounts of the effects of the parameters on the behavior of the model. Some studies indicate that these effects are far from straightforward, see [11] and [10].

In the original volatility equation in (1), the parameter α_0 pushes volatility away from zero and thus imposes an unnecessary restriction on the volatility process. We note by analyzing the series representation in (2) that it attempts to play a dual role: accounting for the location and for the scale simultaneously. In order to address this issue, we propose an extension of the existing model by introducing an additional parameter. This extension is presented first in Section 2 so that the following discussions apply to this generalized model.

The focus of the paper is on analysis of the effect of the six parameters in the volatility equation (1): α_0 , α , β , δ , θ , and the new parameters λ on the features and the so-called ‘stylized facts’ observed in data. We focus, in particular, on the power parameter δ which played an instrumental role when the APARCH model was introduced in [5]. The motivation for introducing δ stemmed from the following argument. Empirical evidence suggested that taking powers of absolute value of returns yields strongest autocorrelation for the case when power is equal to one and thus it appears to support setting $\delta = 1$. However, it was argued, volatility models with different δ (for example $\delta = 2$) may still show similar behavior of absolute powers of returns to the one observed in the empirical data, i.e. the power of one yielding the highest autocorrelation. Therefore there seems to be no an empirical reason to limit the model to the case $\delta = 1$. This provided an argument for introducing δ but its actual effect has not been well investigated. There is some anecdotal evidence about effects of the power parameter for which typical values reported in the literature are somewhere between 0.8 and 2.5, see, for example, [16]. An effect on the tail weight and extent of volatility time dependence (clustering) was claimed. However, a thorough treatment of this is still lacking. Some step in this direction has been made in [10]. Our work extends

these initial findings, Sections 3.1, 4.1, and 4.2.

It is equally important to acknowledge that the range of the parameters has to be established in order for correlations and stationary solutions to be well defined. The initial range of the parameters, $\alpha_0 > 0$, $\alpha \geq 0$, $\beta \in [0, 1)$, $\delta > 0$, $\theta \in [0, 1]$, $\lambda \geq -\beta/(1 - \beta)$ must be complemented by additional restriction to assure that the proposed model for ρ_t yields a well defined stationary process within a space that guarantees the existence of desired moments. For example, it is a common practice to square the data to obtain a proxy for variable variance. If, then, the time dependence is investigated through correlation of the squares, the fourth moment of the volatility is needed, which, we shall demonstrate, imposes certain restrictions on the values of the parameter.

To utilize the model in practice, one has to develop estimation techniques to fit the parameters in the model. The estimation based on the maximum likelihood method under the assumption of normality, where the likelihood is recursively evaluated from the conditional structure of variance, has been extensively used due to its relative simplicity. On the other hand and as reported in our motivating example, the standardized residuals frequently do not follow the assumption of Gaussianity. Thus if one would like to capture the tail behavior, this method does not necessarily accomplish it. Instead, one can estimate kurtosis and match the model kurtosis through the method of moments. The same applies to other stylized facts that can be characterized by moments or functions of them. Finally, one can combine these approaches by matching some sample moments with the ones of interest and maximizing the likelihood under the resulting constraints. This guarantees that the fit model exhibits features important for the problem at hand. All this is discussed in Section 5 and illustrated in Section 6.

2 Scale-location extended volatility model

At the first sight, parameter α_0 in the recurrence relation for volatility in (1) appears to play the role of a location of the model. However, by noticing that the solution to this equation is

$$\rho_t^\delta = \alpha_0 \left(1 + \sum_{k=1}^{\infty} \lambda_{t-1} \dots \lambda_{t-k} \right)$$

one concludes that this parameter rather controls the scale of volatility. In other words, instead introducing α_0 in the equation for the volatility one can equivalently choose e_t to have normal distribution with standard deviation $1/\alpha_0$.

However, and it follows immediately from the above series representation, the volatility process ρ_t as defined can not take values lower than the value of α_0 . In fact, if $\beta > 0$, then this shift from zero extends by the term $\alpha_0\beta/(1 - \beta)$. This separation from zero seems somewhat artificial, since the volatility process should be allowed to take any non-negative value.

We also note that in a certain sense, α_0 does play the role of the location, since it shifts a scaled stationary and non-negative process from zero. Thus the parameter α_0 plays a double role, which somehow limits flexibility of the model. It would be natural to add additional parameter $\lambda \geq -\beta/(1-\beta)$ that will account for an arbitrary shift of the distribution independently of the shift provided by α_0 and β . This is achieved in the following extension that we call the scale-location volatility APARCH model

$$\rho_t^\delta = \alpha_0 \left(\lambda + \sum_{k=1}^{\infty} \lambda_{t-1} \cdots \lambda_{t-k} \right). \quad (2)$$

We note the corresponding recurrent relation

$$\rho_t^\delta = \alpha_0 (\lambda + (1-\lambda)\lambda_{t-1}) + \rho_{t-1}^\delta \lambda_{t-1}, \quad t = \dots, -1, 0, 1, \dots \quad (3)$$

The extended model is equivalent to the original APARCH model for $\lambda = 1$. Another interesting example is the case $\lambda = -\beta/(1-\beta)$ that allows for arbitrary non-zero volatility values.

2.1 Existence and stationarity conditions

We present the conditions on the parameters to guarantee the existence of stationary process ρ_t satisfying recurrence relation of the model. Although it is a mathematical problem, it is also of importance for practitioner when dealing with actual data. In particular, estimated parameters values have to be constrained so it is assured that required covariances or, in general, moments of ρ_t^δ do exist. For example, in [11], it can be seen that for real data the choice of value for δ has to be limited if the covariance of ρ_t^δ is expected to be well-defined. This is further discussed in Subsection 4, where Figure 7 shows the parameter region yielding well defined autocorrelation.

Consider autoregressive volatility model (2), where $\lambda_t \geq 0$ are independent identically distributed non-negative random variables. To provide with conditions for the existence of a stationary volatility process we notice that it is equivalent to the following series being well defined

$$\sum_{k=1}^{\infty} \lambda_{t-1} \cdots \lambda_{t-k}. \quad (4)$$

It should be noted that although mathematically it easier to discuss ρ_t^δ , for the APARCH model the volatility ρ_t is of importance. The question is now in what sense the defining equation yields ρ_t having, for example, the first, the second, or, more generally, the p th moment. The existence is equivalent to convergence of the series $\sum_{k=1}^{\infty} \lambda_{t-1} \cdots \lambda_{t-k}$ in the p/δ norm, which is discussed in detail in the Appendix. We note that once the convergence is established the process ρ_t^δ is strictly stationary, i.e. its finite dimensional distributions are shift invariant. The mathematically most elegant sufficient solutions are obtained for the cases of $p = \delta$ and $p = 2\delta$. We report the corresponding restricting equations for the parameters that follow from Propositions 3 and 4.

The case of $p = \delta$

A sufficient condition for existence of a strictly stationary solution ρ_t^δ with finite first moment is

$$(1 - \theta)^\delta + (1 + \theta)^\delta < \sqrt{\pi} \frac{1 - \beta}{\alpha} \frac{2^{1-\delta/2}}{\Gamma\left(\frac{\delta+1}{2}\right)}. \quad (5)$$

We note two important special cases. Firstly, $\delta = 1$ yields the condition that does not depend on θ :

$$1 < \sqrt{\frac{\pi}{2}} \frac{1 - \beta}{\alpha}. \quad (6)$$

Secondly, the case of $\delta = 2$ yields

$$1 + \theta^2 < \frac{1 - \beta}{\alpha}. \quad (7)$$

The case of $p = 2\delta$

A sufficient condition for existence of a strictly stationary solution ρ_t^δ with finite second moment is

$$2^{\delta/2-1} \Gamma\left(\delta + \frac{1}{2}\right) \alpha \left((1 - \theta)^{2\delta} + (1 + \theta)^{2\delta} \right) + \Gamma\left(\frac{\delta + 1}{2}\right) \beta \left((1 - \theta)^\delta + (1 + \theta)^\delta \right) < \sqrt{\frac{\pi}{2^\delta}} \frac{1 - \beta^2}{\alpha}. \quad (8)$$

Here, we also note two important special cases. Firstly, $\delta = 1$ yields the condition

$$\alpha (1 + \theta^2) + \frac{2^{3/2}}{\sqrt{\pi}} \beta < \frac{1 - \beta^2}{\alpha} \quad (9)$$

Secondly, the case of $\delta = 2$ yields

$$3\alpha(1 + 6\theta^2 + \theta^4) + 2\beta(1 + \theta^2) < \frac{1 - \beta^2}{\alpha}. \quad (10)$$

Remark 1. We note that the above conditions are sufficient and necessary for the absolute convergence in the norm of L_p . These conditions are particularly important if one wants to consider models that have flexibility to account for wide range of values of meaningful parameters such as kurtosis or autocorrelations. For example, to account for large values of kurtosis that relate to the tails of distribution one has to consider the parameters that lie close to the boundary of the region for which the model is defined in L_p for $p = 4$. This effect is illustrated in Figure 4 in the next section. In Figure 2, we see the regions for the parameters α , β and θ that guarantees the existence model for two cases of $\delta = 1$ and $\delta = 2$, that illustrate equations (9) and (10). In the first case we consider the existence in L_2 sense while in the second case the existence in L_4 sense is considered. Both cases are important when the leverage and kurtosis are considered. We also note that the value of λ does not affect the existence conditions.

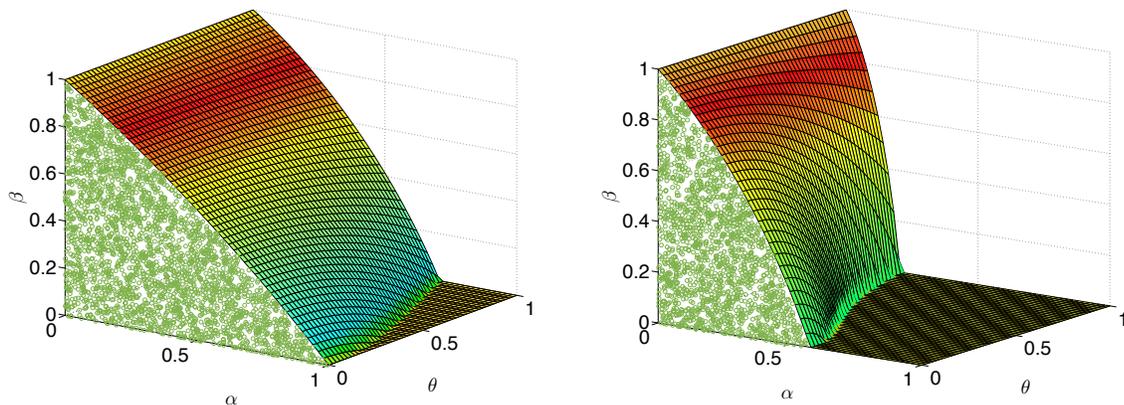


Figure 2: The regions of existence: variance for $\delta = 1$ (left) and kurtosis for $\delta = 2$ (right).

3 Moments and tail behavior

We have seen in the example discussed in the introduction that the model (1) has limited ability to account for the tails of the distribution of the returns. Here we discuss the tail of the distribution in some further detail. We start by noting that although the conditions from the previous section (and more general ones in the Appendix) are aiming at determining when the model is well defined, they also yield information about the tail behavior of ρ_t (and thus of y_t). Namely, if the model ceased to exist in, say, the mean square sense, i.e. the parameters reach the boundary of the region where the moments of λ_t are close to one, then the tails must become heavier eventually yielding infinite variance. Those regions and their boundaries are depending on δ but for the two special cases we have very straightforward relations. The existence of variance of ρ_t is guaranteed by (9), for $\delta = 1$ and by (7) for $\delta = 2$. Consequently tails are heavier when the difference between both sides of the inequalities are approaching zero. This shows that the tail behavior of the model is quite complex and the issue is not particularly well investigated in the literature. In [16], the authors suggest to consider values other than 2 for δ by stating ‘...for non-normal data, by squaring the returns one effectively imposes a structure on the data which may potentially furnish sub-optimal modeling and forecasting performance relative to other power term’. Our results provide mathematical validation of this and similar statements. The simplest way to analyze tail behavior of a random variable is through comparing the moments, the kurtosis being the most popular measure. Here we discuss the relation between moments and the parameters in the model.

We start with parameter δ , which power transforms the volatility and innovation. The reason for introducing this parameter can be confused with the power transform as present in the Box-Cox method, where it is used to transform residuals of the data to fit the tails to normal distribution. Let us clarify here that this is not the case, i.e. the parameter δ is not intended to correct for non-normality of the residuals. Note

that if one observes $y_t = a + \epsilon_t$, then $|\hat{\epsilon}_t|^\delta = |y_t - \bar{y}_t|^\delta$ is the power transformation of the residuals. In the Box-Cox method the reason for considering the power of residuals is their non-normal typically heavy tailed distribution. Thus δ is chosen so that $|\hat{\epsilon}_t|^\delta$ becomes closer to normal distribution (in time series models we would also normalize the residuals given the past). In the APARCH model the situation is opposite. Consider the simpler case of $\theta = 0$ and $a \approx \bar{y}_t$ so that we can set $a = 0$. Then conditionally on the past, the untransformed data $y_t = \rho_t e_t$ are normally distributed while $y_t^\delta = \rho_t^\delta e_t^\delta$ is not. In this sense, the use of δ is exactly opposite to the spirit of using power transformation in the Box-Cox method: one takes power of the data to introduce the volatility relation for otherwise (conditionally) Gaussian data. Let us mention that there is also work in which the time series financial data are first power transformed and then analyzed through the GARCH which in the agreement with the Box-Cox method, see for example [18].

The most direct study of the tail effect of the parameters is to discuss the kurtosis of the returns. Unfortunately, analytical methods of investigating the dependence of kurtosis on δ are limited and one has to resort to numerical methods in the general case. However for two special cases, $\delta = 1$ and $\delta = 2$, the explicit formula are available as shown below. We begin with general formulas for the moments of volatility.

3.1 Moments of volatility

Let M_j be the j th moment of $L = \sum_{k=1}^{\infty} \lambda_{-1} \dots \lambda_{-k}$ and σ_{ij} be the covariances between L^i and L^j . Let us note the following relations that enable to evaluate these parameters by using Lemmas 1 and 2 presented in the appendix

$$\sigma_{ij} = M_{i+j} - M_i M_j. \quad (11)$$

For the two special cases of interest Proposition 6 presented in the appendix allows for more explicit relations. Namely, for the case $\delta = 1$:

$$\mathbb{E}(\rho_t^2) = \alpha_0^2 (\lambda^2 + 2M_1\lambda + M_2) = \alpha_0^2 ((\lambda + M_1)^2 + \sigma_{11}), \quad (12)$$

$$\begin{aligned} \mathbb{E}(\rho_t^4) &= \alpha_0^4 (\lambda^4 + 4\lambda^3 M_1 + 6\lambda^2 M_2 + 4\lambda M_3 + M_4) \\ &= \alpha_0^4 (4\lambda^2 \sigma_{11} + 4\lambda \sigma_{12} + \sigma_{22}) + \mathbb{E}^2(\rho_t^2) \end{aligned} \quad (13)$$

and for $\delta = 2$:

$$\mathbb{E}(\rho_t^2) = \alpha_0^2 (\lambda + M_1), \quad (14)$$

$$\begin{aligned} \mathbb{E}(\rho_t^4) &= \alpha_0^4 (\lambda^2 + 2M_1\lambda + M_2) \\ &= \alpha_0^4 \sigma_{11} + \mathbb{E}^2(\rho_t^2). \end{aligned} \quad (15)$$

Remark 2. The values of M_j and σ_{ij} can be obtained from Lemma 2 of the appendix. For instance, some simple algebra yields the following expression in terms of the mean

m and variance σ^2 of λ_i 's

$$\begin{aligned} E(\rho_t^\delta) &= \alpha_0 \left(\lambda + \frac{m}{1-m} \right), \\ E(\rho_t^{2\delta}) &= \frac{\alpha_0^2}{(1-m)^2} \left(\frac{\sigma^2}{1-\sigma^2-m^2} + (\lambda + m(1-\lambda))^2 \right). \end{aligned}$$

Remark 3. We can use the results with an explicit form of the k th moment m_k of λ_i 's from Lemma 1 of the appendix and their relation with M_j 's listed in Lemma 2 of the Appendix to evaluate the above relations in terms of the actual model parameters. Firstly, for $\delta = 1$, we use the expressions for the first two moments and variance which are

$$\begin{aligned} m &= m_1 = \beta + \alpha \sqrt{\frac{2}{\pi}}, \\ m_2 &= \beta^2 + 2\alpha\beta \sqrt{\frac{2}{\pi}} + \alpha^2(1 + \theta^2), \\ \sigma^2 &= \alpha^2 \left(1 - \frac{2}{\pi} + \theta^2 \right). \end{aligned}$$

Secondly, for $\delta = 2$, the expressions for the mean and variance simplify to

$$\begin{aligned} m &= m_1 = \beta + \alpha(1 + \theta^2), \\ m_2 &= \beta^2 + 2\alpha\beta(1 + \theta^2) + 3\alpha^2(1 + 6\theta^2 + \theta^4), \\ \sigma^2 &= 3\alpha^2(1 + 6\theta^2 + \theta^4) - \alpha^2(1 + \theta^2)^2. \end{aligned} \tag{16}$$

We report below the higher moments which will be required to study the kurtosis. For example, for $\delta = 1$, the expressions for the third and fourth moments of λ_k 's are

$$m_3 = \beta^3 + 3\alpha\beta^2 \sqrt{\frac{2}{\pi}} + 3\alpha^2\beta(1 + \theta^2) + 2\alpha^3(1 + 3\theta^2) \sqrt{\frac{2}{\pi}}, \tag{17}$$

$$m_4 = \beta^4 + 4\alpha\beta^3 \sqrt{\frac{2}{\pi}} + 6\alpha^2\beta^2(1 + \theta^2) + 8\alpha^3\beta(1 + 3\theta^2) \sqrt{\frac{2}{\pi}} + 3\alpha^4(1 + 6\theta^2 + \theta^4). \tag{18}$$

Effect of δ on tails of volatility

It is clear from the above derivations that quantifying the actual effect of δ on the values of the moments is non-trivial since the latter are not explicitly available, except for the special cases shown above. The reason is that, in the main model, the volatility enters as ρ_t and not as ρ_t^δ . Thus to make comparisons between various the case of various values of δ meaningful we need the moments of ρ_t and not of ρ_t^δ . This can be effectively achieved only through numerical studies. We use the Monte Carlo method to investigate the contribution of δ parameter in determining the distributional tails

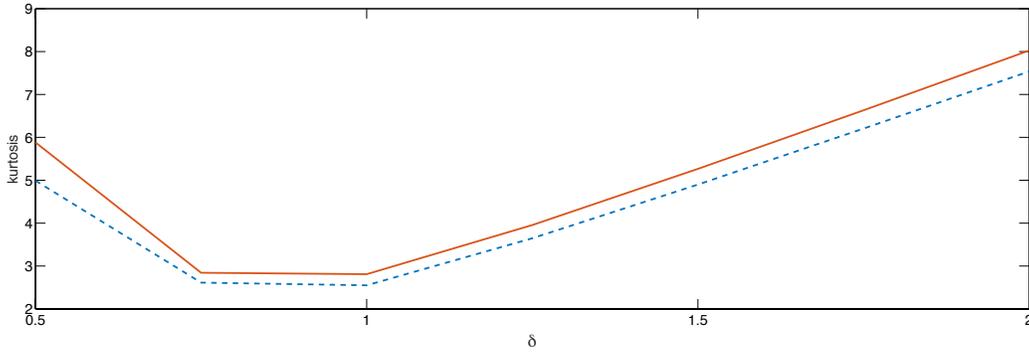


Figure 3: Kurtosis of ρ_t as a function of δ for symmetric ($\theta = 0$ – dashed line) and asymmetric ($\theta = 0.3$ – solid line) cases.

of ρ_t . For this purpose, we evaluate the kurtosis of simulated ρ_t as a function of θ and δ . As an illustration, we take the model that was fit to the S&P 500 data discussed in the introduction and thus we take $\alpha = 0.091$ and $\beta = 0.9$ (these values are also reported in [11]). The findings are reported in Figure 3 for both the asymmetric case with $\theta = 0.3$ taken from the actual fit to the data and the symmetric one ($\theta = 0$), the latter considered for comparison. The presence of asymmetry slightly increases the kurtosis of ρ_t , but the overall pattern is retained. Almost a linear increasing trend can be seen for kurtosis when $\delta \geq 1$.

3.2 Moments of returns

In this section, we derive the variance and kurtosis of returns y_t . Then these are utilized to discuss tail behavior of returns and its deviation from normality. We start with two general relations for the conditionally heteroskedastic model

$$\text{Var}(y_t) = \text{E}(y_t - \text{E}y_t)^2 = \text{E}(e_t^2) = \text{E}(\rho_t^2)\text{E}(e_t^2) = \text{E}(\rho_t^2), \quad (19)$$

$$\kappa_y = \frac{\text{E}(y_t - \text{E}y_t)^4}{\text{Var}^2(y_t)} = \frac{\text{E}(\rho_t^4)\text{E}(e_t^4)}{\text{E}^2(\rho_t^2)} = 3 \cdot \frac{\text{E}(\rho_t^4)}{\text{E}^2(\rho_t^2)} \quad (20)$$

The moments of ρ_t^δ are obtained using Proposition 6, Lemmas 1 and 2 of the Appendix.

Proposition 1. *Let the returns series y_t follows the model in (1) under the efficient market hypothesis, i.e. when the function f does not depend on the past and the conditional variance ρ_t^2 satisfying (2). Let us denote the variance and kurtosis of y_t by σ_y^2 and κ_y respectively and let, as before, M_1 be the mean of $L = \sum_{k=1}^{\infty} \lambda_{-1} \dots \lambda_{-k}$ and σ_{ij} be the covariances between L^i and L^j . We have, for the case of $\delta = 1$:*

$$\begin{aligned} \sigma_y^2 &= \alpha_0^2 \left((\lambda + M_1)^2 + \sigma_{11} \right), \\ \kappa_y &= 3 + 3 \frac{4\lambda^2\sigma_{11} + 4\lambda\sigma_{12} + \sigma_{22}}{\left((\lambda + M_1)^2 + \sigma_{11} \right)^2}. \end{aligned}$$

and for $\delta = 2$,

$$\begin{aligned}\sigma_y^2 &= \alpha_0 (\lambda + M_1), \\ \kappa_y &= 3 + 3 \frac{\sigma_{11}}{(\lambda + M_1)^2}.\end{aligned}$$

Proof. Utilising (19) and (12) for $\delta = 1$, we obtain

$$\text{Var}(y_t) = \alpha_0^2 ((\lambda + M_1)^2 + \sigma_{11})^2$$

and using (14) for $\delta = 2$:

$$\text{Var}(y_t) = \alpha_0 (\lambda + M_1).$$

Similarly utilizing (13) and (20), for the case of $\delta = 1$ we obtain

$$\kappa_y = 3 \cdot \left(1 + \frac{4\lambda^2\sigma_{11} + 4\lambda\sigma_{12} + \sigma_{22}}{((\lambda + M_1)^2 + \sigma_{11})^2} \right)$$

and from (15) it follows that for $\delta = 2$:

$$\kappa_y = 3 \cdot \left(1 + \frac{\sigma_{11}}{(\lambda + M_1)^2} \right).$$

□

Remark 4. It should be stressed that the above formulas for kurtosis require the parameters of the model to lie within the region where the fourth moment of ρ_t (and thus of y_t) is finite. Thanks to our results given in the Appendix, Propositions 3 and 4, we can explicitly identify these regions. This is illustrated in the following example.

Example 1. In Figure 4, we present the graphs of the values of the excess kurtosis evaluated for the model as a function of α and β . We consider δ in $\{1, 2\}$ and λ in $\{0, 1\}$, which leads to four cases. The values of θ is set to zero and the autoregression function f is just a constant. We see that the effect of λ (the new parameter in our extension) is prominent showing that the extension adds flexibility to address features in the data. This model will be discussed in further details in Section 5.

Remark 5. We observe that in both the cases the kurtosis is bigger than in the normal case. Additionally, for $\delta = 1$, by using the Cauchy-Schwartz inequality we achieve the following bounds for the excess kurtosis

$$\frac{3}{\sigma_{11}} \left(\frac{2\lambda - \sqrt{\sigma_{22}/\sigma_{11}}}{1 + (\lambda + M_1)^2/\sigma_{11}} \right)^2 \leq \kappa_e \leq \frac{3}{\sigma_{11}} \left(\frac{2\lambda + \sqrt{\sigma_{22}/\sigma_{11}}}{1 + (\lambda + M_1)^2/\sigma_{11}} \right)^2.$$

Using the relations between M_i 's and σ_{ij} 's given in (11) and Lemmas 1 and 2, one can obtain σ_{11} , σ_{12} and σ_{22} in term of the actual parameters in the model, see also

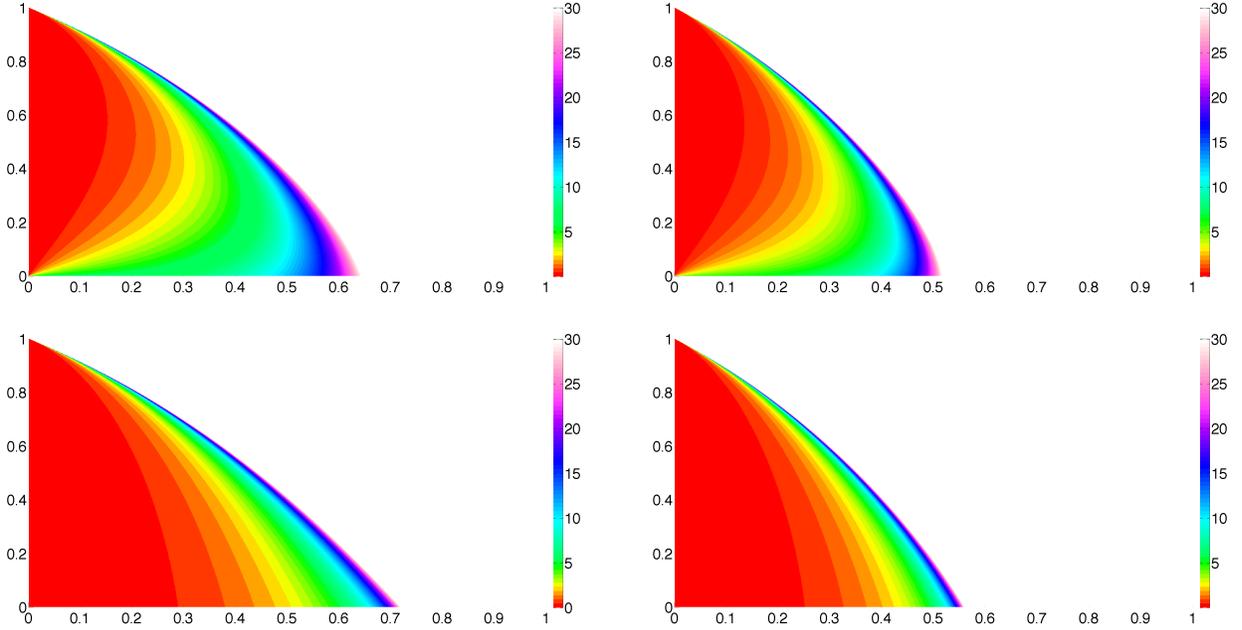


Figure 4: The area for parameters α (horizontal axis) and β (vertical axis) guaranteeing the existence of the kurtosis and heatmap of the excess-kurtosis values: $\delta = 1$ (left) and $\delta = 2$ (right), $\lambda = 0$ (top) and $\lambda = 1$ bottom.

Remark 2 and 3. Thus one can analyze the kurtosis and tails of the returns for a particular specification of the model.

For $\delta = 2$, the excess kurtosis can be rewritten in the terms of the mean m and variance σ^2 of λ_t :

$$\kappa_e = \frac{3}{1 - \sigma^2 - m^2} \frac{\sigma^2}{(\lambda - \lambda m + m)^2}. \quad (21)$$

From this we can notice that the kurtosis increases without bound with $\sigma^2 + m^2$ approaching one, which is the upper bound for existence of the model in the L^4 -sense needed for having kurtosis well defined. Thus, in terms of kurtosis, the distribution of the returns can be made arbitrarily heavy tailed.

4 Dependence structure

The APARCH model follows a symmetric GARCH(1,1) model except it adds two additional parameters θ and δ to account for asymmetric and heavy tail behaviour in the model. Therefore, it is of interest to analyse the contributions of these extra parameters to the autocorrelation function and kurtosis of ρ_t and ϵ_t . The motivation for introducing δ was the Box-Cox power transform which is known to be useful to

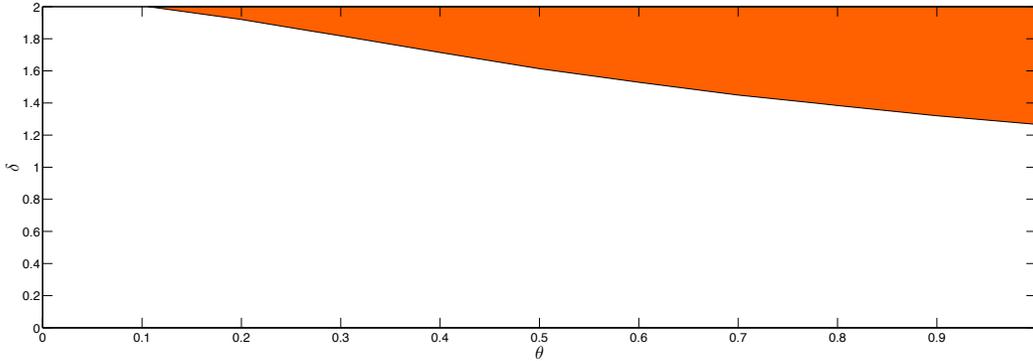


Figure 5: The region of existence of autocorrelation function of ρ_t^δ . The orange area indicates non-convergence in the mean square sense.

reduce anomalies such as non-normality also present in financial data. However, in the model power is not used to transform the data obtained from a certain but rather is part of the model inside of autoregressive structure of ρ_t . Therefore, it is not obvious at all in what way this parameter contributes when it comes to affecting heavy tails (we have seen the complexity of this problem already in the previous subsection). In this subsection, we discuss the role of parameter δ through numerical simulations. However, it is easier mathematically to consider the ρ_t^δ and ϵ_t^δ , which is done next. One has to bear in mind that this is only proxy for the actual effect on ρ_t which is present in the equation for the returns y_t in (1).

4.1 Autocorrelations

The autocorrelation of ρ_t^δ is particularly simple, see Remark 2 in the previous section and Proposition 7 in the appendix,

$$\begin{aligned} r(\rho_t^\delta, \rho_0^\delta) &= m^t, \\ \text{Var}(\rho_t^\delta) &= \frac{\alpha_0^2 \sigma^2}{(1-m)^2(1-\sigma^2-m^2)}. \end{aligned} \tag{22}$$

In the above relations, m and σ^2 are the first moment and the variance of the random variables λ_i .

The necessary and sufficient condition for the existence of the autocorrelation function is that the finite second moment of a considered process must be finite. In our case, for the existence of the autocorrelation of ρ_t^δ , the relation, $E^2(\lambda_t) < 1$, needs to be satisfied. We plot the second moment of ρ_t^δ , $E^2(\lambda_t)$, as a function of θ and δ in order to illustrate how the obtained formulas can assist in determination of the range of parameters for which the model is mathematically meaningful. Figure 5 demonstrates the findings for the model with the parameters other than θ and δ set to the values obtained from the fit to S&P 500 data. The shaded region in the figure is the area of

non-convergence. For a given set of values of θ ranging from 0 to 1 and δ ranging from 0.5 to 2, it can be seen that the convergence cannot always be achievable within the selected bound. It is an important observation and helps in assessing the role of these two extra parameters associated with the long range dependence of the process. It highlights that the autocorrelation exists only for some values within the chosen range.

Let us turn now to the case of $|\epsilon_t|^\delta$. In line with the results of [11] and [1], we have for $t \geq 1$ the following autocorrelation formula

$$r(|\epsilon_t|^\delta, |\epsilon_0|^\delta) = 2 \frac{\sigma^2 m e(\delta) + \alpha \phi(\delta, \theta) (1 - m^2) (e(2\delta) - 2e^2(\delta))}{\sigma^2 e(2\delta) + (\lambda + m(1 - \lambda))^2 (1 - \sigma^2 - m^2) (e(2\delta) - 2e^2(\delta))} e(\delta) m^{t-1}, \quad (23)$$

and for variance we have

$$\text{Var}(|\epsilon_t|^\delta) = 2 \left(\frac{\alpha_0}{1 - m} \right)^2 \frac{\sigma^2 e(2\delta) + (\lambda + m(1 - \lambda))^2 (1 - \sigma^2 - m^2) (e(2\delta) - 2e^2(\delta))}{1 - \sigma^2 - m^2}. \quad (24)$$

For the proofs see Proposition 8, in the appendix. There we use $\nu_p = 2e(p)$, where $e(p)$ is given in (36), and $\gamma(\delta) = \text{Cov}(\lambda_0, |e_0|^\delta) = 2\alpha \phi(\delta, \theta) (e(2\delta) - 2e^2(\delta))$, where $\phi(\delta, \theta) = (1 - \theta)^\delta + (1 + \theta)^\delta$. To get the formulas in terms of the model parameters one can utilize Lemma 1 in the appendix.

For illustration we present in Figure 6 the lag-one autocorrelations for $|\epsilon_t|^\delta$ and ρ_t^δ that are obtained using the presented formulas. We can see that generally an increase in δ increases the autocorrelation (specifically for $\delta \geq 1$). This effect is more dramatic for the autocorrelation of $|\epsilon_t|^\delta$ than that of ρ_t^δ . Moreover the asymmetry has somewhat large effect on the autocorrelation of $|\epsilon_t|^\delta$ for δ close to 2.

The effect of δ on the long-term memory

The above results on the dependence in the model can not be indicative of the actual effect of δ . In fact, these theoretical formulas for autocorrelation functions, although explicit, does not give us correlations in y_t , except for the case $\delta = 1$. Therefore we do not discuss any further the effect of δ on the computed correlations of powers of ρ_t , while more discussion can be found in [11]. Instead, we move on to numerically assess the role of the parameter δ on the autocorrelation of ρ_t in the case of model fit to the S&P 500 data.

Figure 7 shows the lag-one autocorrelation of ρ_t and $|\epsilon_t|$ for the symmetric case ($\theta = 0$) and an asymmetric case ($\theta = 0.3$). It can be seen, from the lower panel of the figure, that the autocorrelation for ρ_t does not change much in these two cases so it appears that the parameter θ does not contribute significantly to the long-memory of ρ_t . Since the process ρ_t depends on its past values (ρ_{t-1}), we see values of the autocorrelation in proximity of one. Moreover, the dependence becomes stronger for

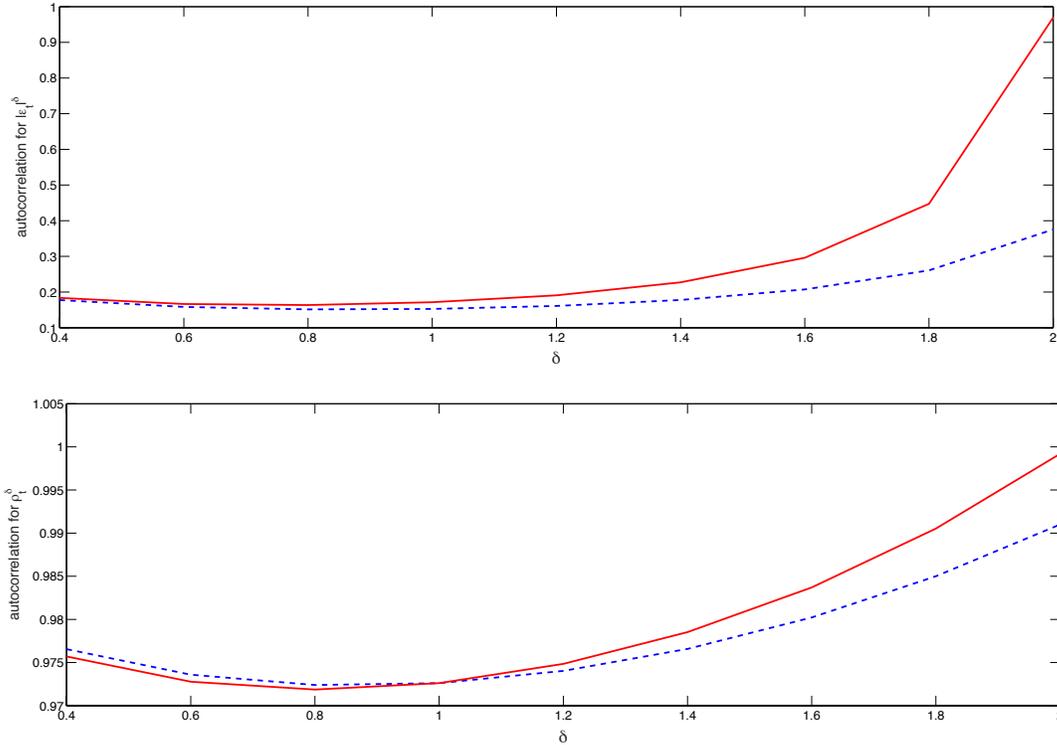


Figure 6: First-order autocorrelation function for $|\epsilon_t|^\delta$ (top) and ρ_t^δ (bottom). Dashed line corresponds to the symmetric case of $\theta = 0$ and the solid line show the case of $\theta = 0.3$.

$\delta \geq 1$. The upper panel of the figure displays the autocorrelation pattern for $|\epsilon_t|$. Here again the parameter θ does not contribute to the long memory as much as in the case of $|\epsilon_t|^\delta$. Moreover, the autocorrelation for $|\epsilon_t|$ is an increasing function of δ , though the dependence is no longer so strong in magnitude, when δ is close to 2, reaching only approximately 0.6, while in Figure 6 the corresponding value is nearly one in the asymmetric case.

4.2 The leverage effect

In general terms, the leverage effect is described as the higher volatility after ‘bad news’ stretches as compared with the volatility during ‘good news’ periods. If bad news are revealed by negative log-return values, then one can measure this effect by correlation between the return y_{t-1} and the volatility ρ_t (see, for example [21]). Generally, negative value of such correlation indicates existence of the leverage effect and the larger the absolute value of the correlation the stronger leverage effect. Here we present some explicit formulas for relevant correlations in the APARCH model that allow to analyze which of the parameters influence the leverage. These formulas can be used for evaluation of the strength of the leverage effect or can be utilized in estimation as discussed

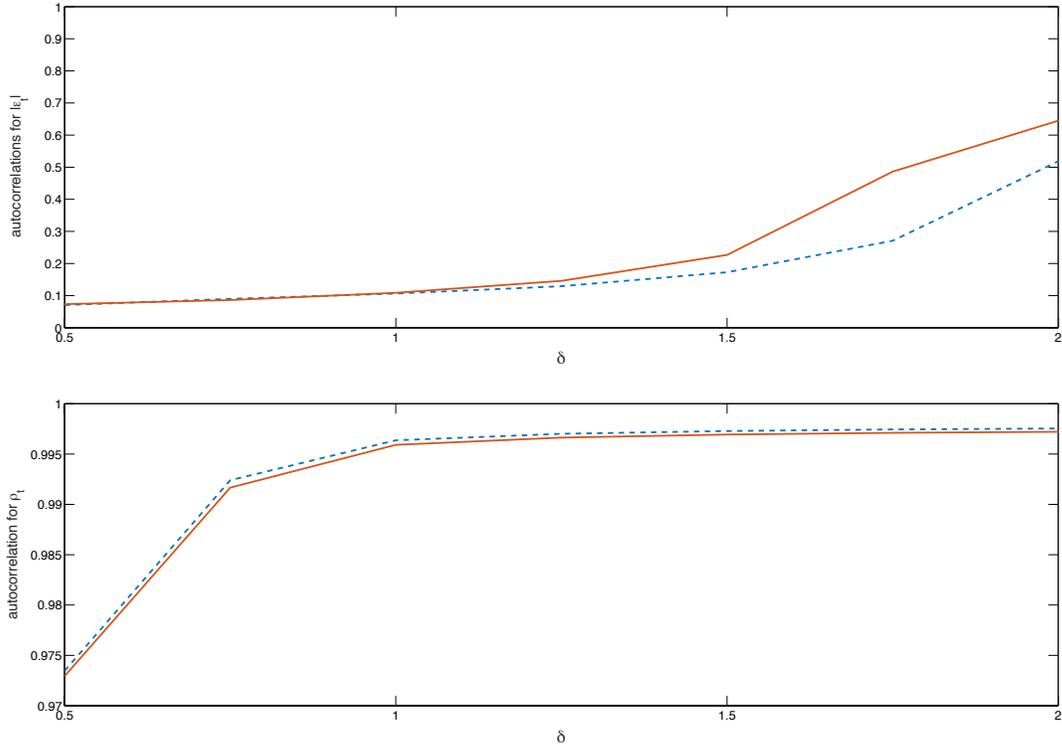


Figure 7: Autocorrelation function for $|\epsilon_t|$ (top) and for ρ_t (bottom) in the symmetric case of $\theta = 0$ (dashed line) and for the case of $\theta = 0.3$ (solid line).

in Section 5.

For a random variable e_t having a symmetric distribution around zero, a proxy of leverage effect is defined through the correlation between lagged δ -powers of centered returns ($\epsilon_{t-1}^{(\delta)}$) and δ -powers of volatility (ρ_t^δ). Here for a number x we define $x^{(\delta)} = x^{+\delta} - x^{-\delta}$. It is shown in the Appendix in Proposition 9 that

$$r(\rho_t^\delta, \epsilon_{t-1}^{(\delta)}) = \alpha \sqrt{e(2\delta)(1 + cv_1^{-2})} \frac{(1 - \theta)^\delta - (1 + \theta)^\delta}{\sqrt{2}}, \quad (25)$$

where $e(p)$ can be simply computed from (36) and the coefficient of variation cv_1 for ρ_0^δ that can be computed explicitly from (22) and Remark 2 is yielding

$$cv_1^{-2} = \frac{1 - \sigma^2 - m^2}{\sigma^2} (\lambda(1 - m) + m)^2, \quad (26)$$

where Corollary 2 in the appendix can be used to obtain explicit forms for m and σ^2 (the mean and variance of λ_i 's) in terms of α , β , δ , and θ .

Two special cases are of particular interest due to their simplicity. For compactness

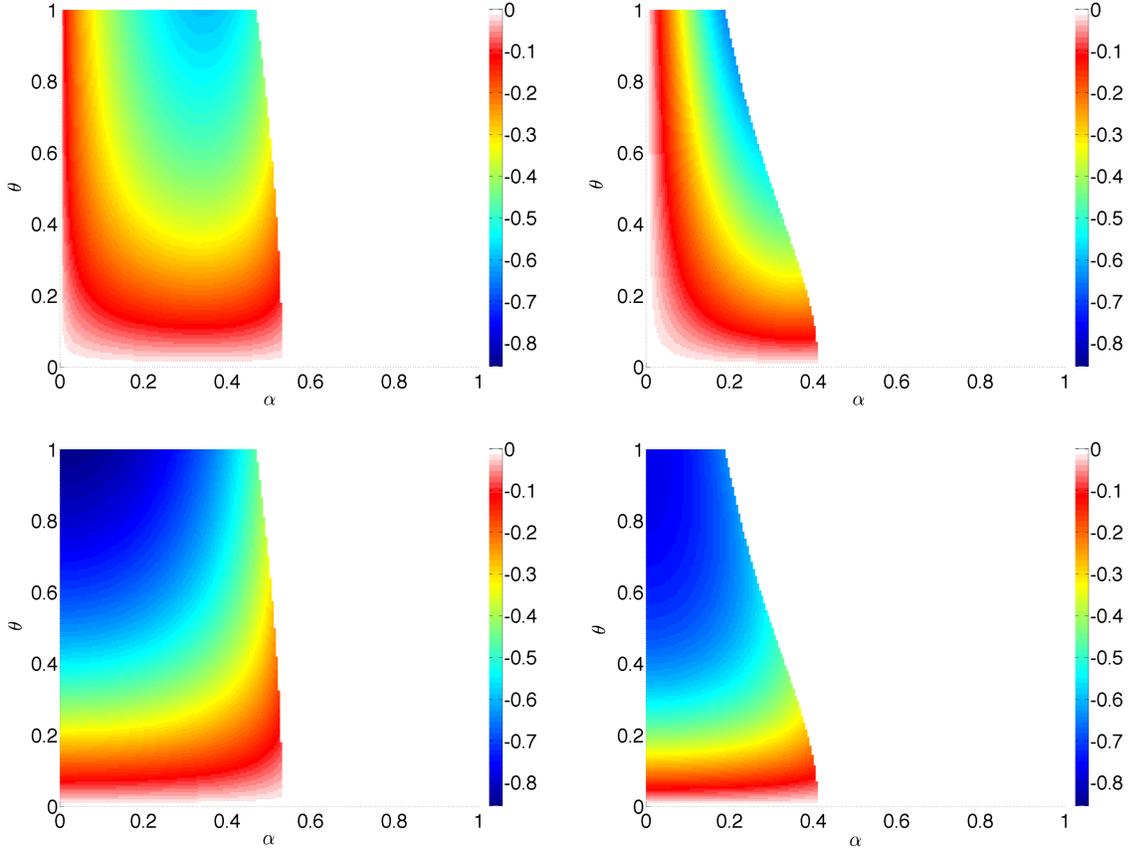


Figure 8: The area for parameters α (horizontal axis) and θ (vertical axis) displaying the heatmap of the leverage effect values: $\delta = 1$ (left) and $\delta = 2$ (right), $\lambda = 0$ (top) and $\lambda = 1$ bottom.

we set $c_{\lambda m} = \lambda(1 - m) + m$. For the case of $\delta = 1$:

$$\begin{aligned}
 r(\rho_t, \epsilon_{t-1}) &= -\alpha\theta\sqrt{1 + cv_1^{-2}} \\
 &= -\frac{\theta}{\sqrt{1 - \frac{2}{\pi} + \theta^2}} \left(\alpha^2 \left(1 + \theta^2 - \frac{2}{\pi} \right) (1 - c_{\lambda m}^2) + (1 - m^2)c_{\lambda m}^2 \right)^{1/2}.
 \end{aligned} \tag{27}$$

Similarly, for $\delta = 2$:

$$\begin{aligned}
 r(\rho_t^2, \epsilon_{t-1}^{(2)}) &= -2\sqrt{3}\alpha\theta\sqrt{1 + cv_1^{-2}} \\
 &= -\frac{\sqrt{6}\theta}{\sqrt{1 + 8\theta^2 + \theta^4}} (2\alpha^2 (1 + 8\theta^2 + \theta^4) (1 - c_{\lambda m}^2) + (1 - m^2)c_{\lambda m}^2)^{1/2}.
 \end{aligned} \tag{28}$$

From the above formulation, it is obvious that the correlation is always negative, if θ is positive, assuring negative association between lagged return and volatility. The fact

that an increase of either α or θ amplifies the effect is not surprising. However, it is interesting to note that the leverage can also increase from a decrease of the coefficient of variation for volatility. For example, the location parameter λ , presented in our extended model (2), enters through the coefficient of variation (cv_1) of volatility in (26). It is interesting to notice that the higher the value of λ , the stronger it will amplify the leverage effect in the data. This property will be revisited in Section 5 where we discuss estimation strategies. Finally, we notice that for existence of the model it has to be assumed that m and $m^2 + \sigma^2$ are bounded by one thus these two terms have limited effect on the leverage.

Example 2. To illustrate the dependence of the leverage on the parameters we consider the case of $\alpha = \beta$ and $\lambda = 1$ (autoregressive function f is constant as always in our examples). In Figure 8 we see the dependence of the leverage as defined through (27) and (28) for the cases $\delta = 1$ and $\delta = 2$, respectively.

Let us also notice the following simplified formulas for the case $\delta = 1$ and $\delta = 2$ for the correlation of the lagged returns and the squared volatility, which is also often used as the measure of the leverage effect. For the case of $\delta = 1$ we have

$$r(\rho_t^2, \epsilon_{t-1}) = -2\alpha\theta\sqrt{\frac{1 + cv_1^2}{cv_2^2}} \left(\alpha_0 + \left(2\sqrt{\frac{2}{\pi}}\alpha + \beta \right) \frac{E(\rho_0^3)}{E(\rho_0^2)} \right) \frac{E(\rho_0)}{E(\rho_0^2)},$$

and for $\delta = 2$:

$$r(\rho_t^4, \epsilon_{t-1}^{(2)}) = -8\sqrt{3}\alpha\theta\sqrt{\frac{1 + cv_1^2}{cv_2^2}} \left(\alpha_0 + (15\alpha(1 + \theta^2) + \beta) \frac{E(\rho_0^6)}{E(\rho_0^4)} \right) \frac{E(\rho_0^2)}{E(\rho_0^4)}.$$

We conclude this section with a general relation that relates the correlation of the powers of the absolute returns with the lagged returns. We can see that they are closely related to the above correlations. This fact can be utilized to compute method of moments estimators that would match the leverage observed in the data. Namely, for each $k > 0$ and for APARCH models such that the correlations below are well defined, we have

$$r(|\epsilon_t|^{k\delta}, \epsilon_{t-1}^{(\delta)}) = r(|e_t|^{k\delta} \rho_t^{k\delta}, \epsilon_{t-1}^{(\delta)}) = 2e(k\delta)r(\rho_t^{k\delta}, \epsilon_{t-1}^{(\delta)}),$$

where the last equality follows from the independence of e_t from e_{t-1} and ρ_t .

5 Estimation strategies

Before getting into details of estimation for the APARCH model let us briefly recap its structure and the role of parameters. We consider the extended APARCH model with the additional parameter λ that is given through

$$\begin{aligned} y_t &= g(\mathbf{y}_{t-1}; \mathbf{a}) + \epsilon_t, & \epsilon_t &= \rho_t e_t \\ \rho_t^\delta &= \alpha_0 (\lambda + (1 - \lambda)\lambda_{t-1}) + \rho_{t-1}^\delta \lambda_{t-1}, \end{aligned} \tag{29}$$

where, $\mathbf{y}_t = (y_s; s \leq t)$, $\lambda_t = \alpha [(1 - \theta)^\delta e_t^{+\delta} + (1 + \theta)^\delta e_t^{-\delta}] + \beta$. Here unspecified time dependent part of the main equation $g(y_{t-1}, y_{t-2}, \dots; \mathbf{a})$ is controlled by the multivariate parameter \mathbf{a} . As it will be discussed, the function g and thus also the parameter \mathbf{a} does not change estimation strategy in any other way than just by accounting the space for \mathbf{a} additionally to the space of other parameters over which likelihood function is maximized. In the simplest but important case g is a linear autoregressive model in which \mathbf{a} represents the underlying regression coefficients. As we have mentioned before, the special case of g being a constant function is in the literature referred to as the random walk model based on the efficient market hypothesis. The remaining parameters are more pertained to empirically observed ‘stylized facts’ that the APARCH model aims to capture. This relation between ‘stylized facts’ and the corresponding parameters is discussed next.

We have introduced the extended volatility model that adds important flexibility which was missing in the original formulation. This extension is characterized by two parameters α_0 and λ that account both for the scale and the location shift in the volatility equation. The role of these parameters has been discussed already in detail in Section 2. They join α and β to fully describe the time dependence in the volatility (the ‘AR part’ of APARCH). Additional two parameters, θ and δ are defining the asymmetric power structure (the ‘A-P part’ of APARCH). We note that the role of θ as a leverage effect parameter has been confirmed in this work, while the role of δ is somewhat ambiguous.

In the full formulation the model is defined by the multivariate parameter

$$\boldsymbol{\theta}^T = [\mathbf{a} \quad \alpha_0 \quad \alpha \quad \beta \quad \lambda \quad \theta \quad \delta]$$

and, when conditioned on the past, features Gaussian likelihood with varying variance (heteroscedasticity) represented by ρ_t (the ‘CH part’ of APARCH). This is the key property that enables a numerically effective maximum likelihood procedure leading to an estimate of $\boldsymbol{\theta}$. The resulting estimate can be viewed as a standard maximum likelihood estimate (nevertheless, some restrictions on parameters following from the model existence conditions have to be imposed).

The maximum likelihood estimators have many well-studied theoretical advantages but these benefits are valid if one truly believes that the mechanism producing real data indeed follows the assumed structural and distributional model. However, in practice, real data rarely strictly follows the model in all its features as it is implicitly assumed by the likelihood method. The likelihood method applied to such data is often referred to as the quasi-likelihood method. It can lead to a statistical fit that does not represent well such ‘stylized facts’ as leverage, heavy tails, or dependence structure. This was the case in our motivating example, where the tails were inaccurately fit despite the model having the capacity of yielding heavy tails (high kurtosis) as shown in Section 3. One may argue that the data may have some features that drives the likelihood estimates seem away from the observed ‘stylized facts’.

As a remedy to this problem a practitioner may maximize likelihood while preserv-

ing certain empirical characteristics describing stylized facts. These restrictions reduce dimensionality of the parameter space so that maximizing of the likelihood function is carried over less dimensional space of parameters. There are two benefits of this approach. Firstly, the important features as seen in the data are followed closely in the estimated model. Secondly, by reducing dimensionality of the optimization problem, the computational cost is reduced.

While the computational benefit is practically justified by numerical limitations of MLE method, following the empirical characteristics of the data is more of the methodological nature. Namely, real data seldom follows rather simplistic models and therefore relying only on the likelihood may not account on the features of interest. Assisting the likelihood through the restrictions make the estimation more robust on the deviations of the data from the model. This is further discussed in Subsection 5.2.

In the remaining parts of this section we provide technical details of both the standard and constrained maximum likelihood estimation methods and discuss the impact the imposed restriction have on the efficiency of estimation.

5.1 The likelihood method

The discussed volatility model is based on conditional heteroscedasticity of innovations and its explicit Gaussian likelihood. The standard likelihood estimation method is simply based on the maximizing likelihoods, where the likelihood is recursively evaluated from the conditional structure of the variance. Here we discuss briefly this estimation technique in some further detail.

Let us represent the likelihood conditionally on the initial past \mathbf{e}_0 through the product rule

$$L(\boldsymbol{\theta}; y_1, \dots, y_t | \mathbf{e}_0) = f_{\boldsymbol{\theta}}(y_1, \dots, y_t | \mathbf{e}_0) = f_{\boldsymbol{\theta}}(y_t | \mathbf{y}_{t-1}, \mathbf{e}_0) \cdots f_{\boldsymbol{\theta}}(y_1 | \mathbf{e}_0),$$

where where $\mathbf{y}_{t-1} = (y_s; s < t)$ and $\mathbf{e}_0 = (e_s; s \leq 0)$ and we assume that the initial history \mathbf{e}_0 is available. In practice, one chooses some initial guesses for \mathbf{e}_0 values using some properties of the model. For example one can take random sample from the distribution of \mathbf{e}_0 . Typically in large sample case, the accuracy of the estimation does not suffer by such a substitution of the true values by their ‘educated’ guesses.

Specifically, the APARCH model can be written in the following form

$$y_t = g(\mathbf{y}_{t-1}, \mathbf{a}) + \rho(y_{t-1}, \dots, y_1, \mathbf{e}_0, \mathbf{b}) e_t \quad (30)$$

where \mathbf{a} and $\mathbf{b} = (\alpha_0, \alpha, \beta, \lambda, \theta, \delta)$ are vectors of parameters. Using the standard normality of e_t ’s, the log-likelihood that we want to maximize with respect to $\boldsymbol{\theta} = (\mathbf{a}, \mathbf{b})$ takes the form

$$l(\boldsymbol{\theta}; y_1, \dots, y_t | \mathbf{e}_0) = -\frac{t}{2} \log(2\pi) - \frac{1}{2} \sum_{k=1}^t \left(\frac{y_k - g(\mathbf{y}_{k-1}, \mathbf{a})}{\rho(y_{k-1}, \dots, y_1, \mathbf{e}_0, \mathbf{b})} \right)^2 - \sum_{k=1}^t \log \rho(y_{k-1}, \dots, y_1, \mathbf{e}_0, \mathbf{b}). \quad (31)$$

Let us consider even a more specific model with $\boldsymbol{\theta} = (\mu, a, \alpha_0, \alpha, \theta, \beta, \delta)$ and

$$\begin{aligned} g(\mathbf{y}_{t-1}, (\mu, a)) &= \mu + ay_{t-1} \\ \rho_t &\stackrel{def}{=} \rho(\mathbf{y}_{t-1}, (\alpha_0, \alpha, \theta, \beta, \delta)) \end{aligned} \quad (32)$$

is given through the recursive relation

$$\rho_t^\delta = \alpha_0 + \alpha_1 \rho_{t-1}^\delta [(1 - \theta)^\delta e_{t-1}^{+\delta} + (1 + \theta)^\delta e_{t-1}^{-\delta} + \beta],$$

or through the non-recursive series representation (2) in the Appendix. The log-likelihood function takes the form

$$l(\boldsymbol{\theta}; y_1, \dots, y_t | y_0) = -\frac{t}{2} \log(2\pi) - \sum_{k=1}^t \log \rho_k - \sum_{k=1}^t \frac{(y_k - \mu - ay_{k-1})^2}{2\rho_k^2}. \quad (33)$$

We note that ρ_t given in (32) is in fact a function of parameters and past observations \mathbf{y}_{t-1} , while, in practice, the entire past \mathbf{y}_{t-1} is not known. So to evaluate the log-likelihood for given values of the parameters one has to provide with the initial values for (y_0, e_0, ρ_0) and evaluate all ρ_k 's, $k = 1, \dots, t$ using the recursive relations (1) and observed (y_1, \dots, y_t) . There are several 'educated' ways of choosing initial values. For example e_0 can be simulated, ρ_0 can be taken as the standard deviation of the data s_y due to (19), and $y_0 = \bar{y}$. Alternatively, the explicit moments formula can be utilized for the particular choice of parameters over which the likelihood will be maximized. For large data sizes this initial choice will not have a significant effect for the final maximizer and thus the issue is not discussed any further.

Example 3 (Maximum likelihood for the APARCH model without leverage). To illustrate how the MLE can be facilitated for APARCH model in practice, we consider (1) with the autoregressive part being constant $f(\mathbf{y}) = \mu$ and theta $\theta = 0$ (no leverage). We first consider $\delta = \lambda = 1$. For particular values of the model parameters, we can recursively evaluate the log-likelihood as described by, for example, starting with the values of $y_0 = \bar{y}$, $\rho_0 = s_y$, and $e_0 = 0$

$$\rho_1 = \alpha_0 \left(1 + \sum_{k=1}^{\infty} \lambda_0 \lambda_{-1} \dots \lambda_{-k+1} \right)$$

and thus for $i \geq 1$:

$$e_i = \frac{y_i - \mu}{\rho_i}, \quad \rho_{i+1} = \alpha_0 + \rho_i \lambda_i,$$

where $\lambda_i = \alpha |e_i| + \beta$ yielding $\rho_1, \dots, \rho_{t+1}$.

Similarly, for $\delta = 2$ we take

$$\rho_1 = \alpha_0^{1/2} \left(1 + \sum_{k=1}^{\infty} \lambda_0 \lambda_{-1} \dots \lambda_{-k+1} \right)^{1/2}$$

and $\rho_{i+1} = \sqrt{\alpha_0 + \rho_i^2 \lambda_i}$ where $\lambda_i = \alpha e_i^2 + \beta$.

The log-likelihood becomes

$$l(\mu, \alpha_0, \alpha, \beta; y_1, \dots, y_t | \mathbf{e}_0) = -\frac{t}{2} \log(2\pi) - \frac{1}{2} \sum_{k=1}^t \left(\frac{y_k - \mu}{\rho_k} \right)^2 - \sum_{k=1}^t \log \rho_k. \quad (34)$$

In Figure 9, the log-likelihood function is presented as a function of α and β (for illustration purposes α_0 is set to one and $\mu = 0$) for $t = 5000$ data simulated from the model. Two pairs of (α, β) have been selected. The first pair, $\alpha = 0.12$ and $\beta = 0.86$, has values close to the one obtained in estimation for the S&P 500 data analyzed in Section 6. This model has a moderate kurtosis value of 3.96, when $\delta = 1$ and rather high kurtosis of 11 for $\delta = 2$, as can be seen from Figure 4 (*bottom*), see also Proposition 1. This choice of parameters places them very close to the boundary of the region guaranteeing existence of the fourth moment (and thus of the kurtosis). The effect of high values of the kurtosis is discussed in the next subsection. The second pair, $\alpha = 0.2$ and $\beta = 0.6$, represents values that are further from the boundary and thus faring fairly small kurtoses (3.51 for $\delta = 1$ and 3.86 for $\delta = 2$), which are closer to the value three featured by the Gaussian distribution. In this figure the likelihood function is shown together with the contour line representing the constrained likelihood method described in the next subsection. We see that the likelihood method retrieves reasonably well the values of the parameters for both cases of δ .

In the above example, it is illustrated how important for the model fitting is to know the range of parameter guaranteeing the existence of the model in a proper mathematical sense. In practice, observing values the MLE close to the boundary of existence of certain moments can influence on the choice of the characteristics used to study the model. For example, by evaluating the likelihood function over the region of finite kurtosis, one can assess how reasonable is the assumption of the finite kurtosis – the MLE close to the boundary may indicate that the data may require releasing this assumption. Even if one requires the existence of certain moments for the model, their values can be used to obtain more realistic models. This is described in the next section.

5.2 The constrained likelihood method

It follows from the general theory of statistics that in most cases the maximum likelihood method provides efficient estimators. We have seen in the previous section that the method can be successfully applied to the extended AP-GARCH model. However, there are several reasons for which one can consider a modification of the method. Some of them have been already discussed in the introduction to this section. Here we reemphasize two that are particularly important in the context of estimation for the extended AP-GARCH model.

Real data very rarely truly follow the model one tries to fit. The MLE in such contexts is called the quasi-likelihood method and still provides consistent estimates

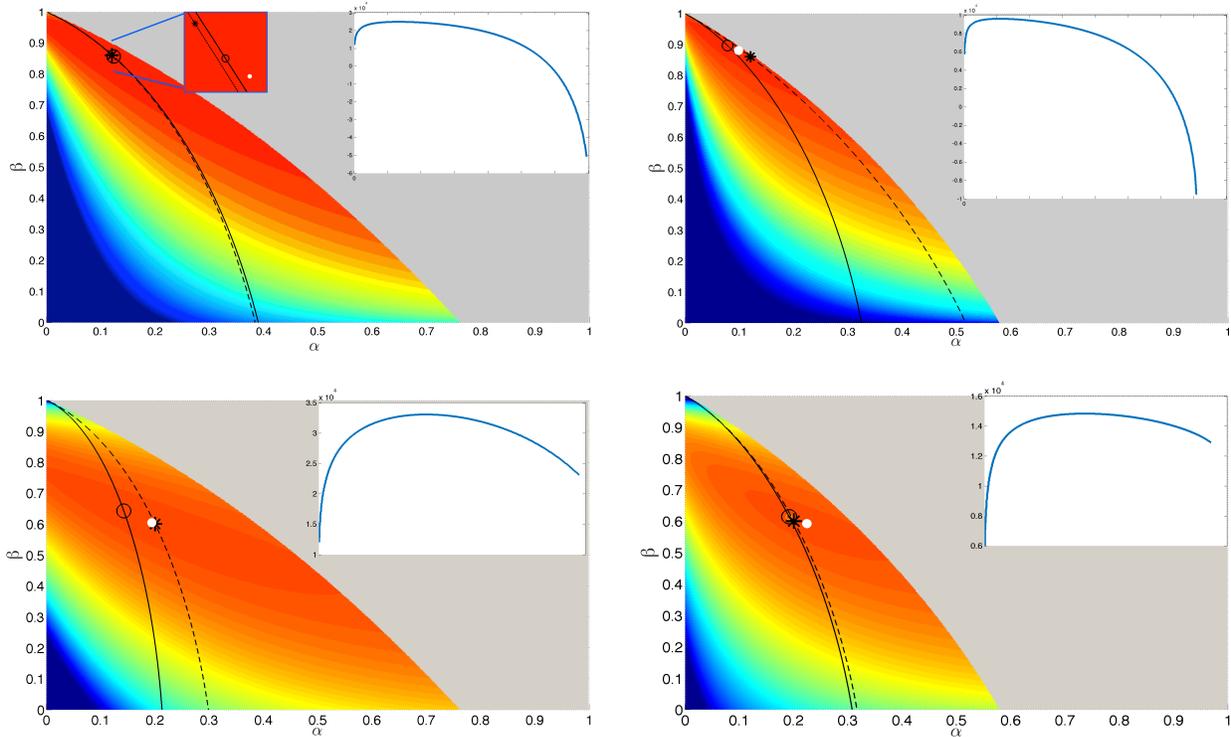


Figure 9: Illustration of the standard and constrained likelihood methods for estimation of α and β for the model in Example 3. The cases of $\delta = 1$ and $\delta = 2$ are presented in the left and the right columns, respectively. The data were simulated for two sets of the parameters: the top graphs feature values $\alpha = 0.12$ and $\beta = 0.86$ that are close to the boundary of the kurtosis existence region, the bottom ones correspond to $\alpha = 0.2$ and $\beta = 0.6$. The true values are represented by the stars on all figures (in the top-left we included an enlargement in order to distinguish between other points on the graph). The gray area represents prohibited region for the parameters (the fourth moment does not exist). The log-likelihood is represented by the colored contour map. The MLE estimator is presented by the white points. The continuous line represents the constrained line so that the parameters match the empirical kurtosis. The dark dot lying on it represents the constrained likelihood estimator. The dashed lines show the constrained log-likelihood line corresponding to the true kurtosis of the model. In the upper left corner of each graph the log-likelihood values along the constraint curve of the constant empirical based kurtosis are shown.

which are relatively easy to evaluate. However, for the model that may not necessarily represent given data in all its features, the quasi-MLE fit may emphasize those features that are not important for a particular application. It can be then desirable to use the likelihood but at the same time to require that some characteristics observed in the data are preserved by the fit to the model.

Secondly, the MLE estimator based on the data can lead to the values of the

parameters that prevents existence of certain characteristics that are important for the model. For example, the estimated parameters can go beyond the region where the kurtosis is well defined. This can happen even if the data are generated from the model with a well defined kurtosis, in particular, when the true parameter values are close to the boundary of kurtosis existence region (the cases of large kurtoses). This problem maybe amplified by the problem mentioned in the previous paragraph.

A simple and quite effective approach to eliminate the above problems can be based on blending maximizing likelihood with restriction that makes values of empirical characteristics to match theoretical characteristics represented by the parameters in the model. Such a ‘hybrid’ estimation scheme can be described next.

Here we use the notation of the previous subsection. Let $S_i(\boldsymbol{\theta})$, $i = 1, \dots, p$ denote certain characteristics of the model as function of the parameter $\boldsymbol{\theta}$. Important examples of such characteristics for financial applications are kurtosis and measures of leverage. Assume that some direct estimators of these characteristics are given as $\hat{S}_i(y_1, \dots, y_t)$, $i = 1, \dots, p$, where y_1, \dots, y_t are the observed values from the model. The constrained MLE $\hat{\boldsymbol{\theta}}$ is defined as the solution to the following optimization problem with constraints

$$\hat{\boldsymbol{\theta}} = \operatorname{argmax}\{l(\boldsymbol{\theta}; y_1, \dots, y_t | \mathbf{y}_0) : \boldsymbol{\theta} \in \Theta, S_i(\boldsymbol{\theta}) = \hat{S}_i(y_1, \dots, y_t), i = 1, \dots, p\}.$$

In the following example we use the kurtosis based constraint that seems to perform quite well.

Example 4 (Kurtosis constrained maximum likelihood). We continue to work with the model from Example 3. For the likelihood we take $l(\alpha, \beta; y_1, \dots, y_t | \mathbf{e}_0)$ given in (34), with $\alpha_0 = 1$ and $\mu = 0$. We provide the formulas only for the case of $\delta = 2$ although the case of $\delta = 1$ can be treated similarly (the formulas are slightly more complex). The kurtosis in this special case is given by

$$\kappa(\alpha, \beta) = 3 \frac{1 - (\alpha + \beta)^2}{1 - (\alpha + \beta)^2 - 2\alpha^2}.$$

Consequently, the problem reduces to finding maximum of $l(\alpha, \beta; y_1, \dots, y_t | \mathbf{e}_0)$ with respect to α and β assuming additionally that the following constraint is satisfied

$$3 \frac{1 - (\alpha + \beta)^2}{1 - (\alpha + \beta)^2 - 2\alpha^2} = \frac{\overline{(y - \bar{y})^4}}{\left(\overline{(y - \bar{y})^2}\right)^2}.$$

This constraint reduce the problem of finding maximum over one parameter (along the curve). In Figure 9, we see the contour line corresponding to this constraint for four cases discussed in Example 3. The log-likelihood values along this contour line are presented in the upper right corner of the graphs. We observe that the method leads to the estimates that are comparable to the one obtained through standard likelihoods.

A choice of the type of constraint can be dictated by its importance in a particular application, its accessibility for direct estimation, and finally by its properties. For

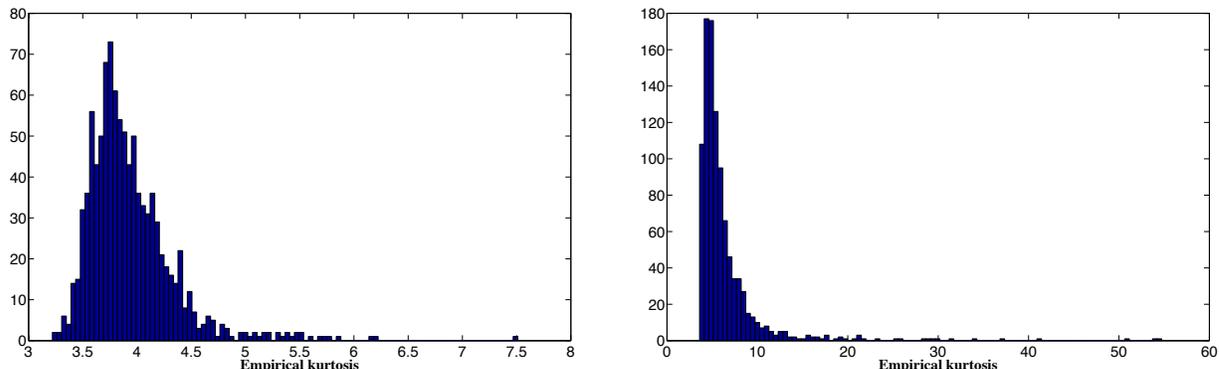


Figure 10: Distribution of empirical kurtosis based on 1000 Monte Carlo samples of size 5000 for the cases discussed in Examples 3 and 4. The parameters are $\alpha = 0.12$ and $\beta = 0.86$. The case of $\delta = 1$ is on the left hand side graph – the theoretical kurtosis is 3.96 and the average of MC sample kurtoses is 3.9374. The case of $\delta = 2$ is on the right hand side graph – the theoretical kurtosis is 11 and the average of MC sample kurtoses is 6.4634. The pronounced bias of the empirical kurtosis for large values of the kurtosis is evident. One observes also the skewness of the sample kurtosis distribution in both the cases.

example, one can choose kurtosis as an important characteristic that one wishes to preserve in the estimation procedure. Kurtosis is easy to estimate by the method of moments and often is used to compare models. But the word of caution is needed here since convenience is not always the best guide in the choice of constraints. This is due to the fact that poor properties of \hat{S}_i can hamper the effectiveness of the method. Here we illustrate this problem by showing the behavior of the sample kurtosis for the AP-GARCH model. In Figure 10 (*right*), we observe that the sample kurtosis for large value of the kurtosis is negatively biased and has distribution that is heavily skewed to the right. The bias is not present for the small values of kurtosis Figure 10 (*left*), but skewness of the distribution still is quite visible. It implies that using sample kurtosis for the constraints may lead to some inaccuracies in the estimation. However, in the previous example using kurtosis did not lead to any apparent problems though more studies are needed to understand the effect of the constraints.

We conclude this section with an explicit form of the kurtosis constrained maximum likelihood method in the full model with $\delta = 2$.

Kurtosis constraint in the full model, the case of $\delta = 2$

The formulas derived in our work allows for evaluation of the kurtosis constraint for the full model in the case of $\delta = 2$. Similar approach can be applied for $\delta = 1$ although the formulas would be more complex. We note if $\delta = 2$, then from (21) and (16) we

obtain the explicit form for the kurtosis

$$\kappa = 3 + 3 \frac{\sigma^2}{1 - m_2} \frac{1}{(\lambda(1 - m_1) + m_1)^2},$$

where

$$\begin{aligned} m_1 &= \alpha (1 + \theta^2) + \beta, \\ m_2 &= 3\alpha^2 (1 + 6\theta^2 + \theta^4) + 2\alpha\beta (1 + \theta^2) + \beta^2, \\ \sigma^2 &= 2\alpha^2 (1 + 8\theta^2 + \theta^4) \end{aligned}$$

and to have the model well defined and possessing all required moments the following range for the parameter is imposed

$$-1 \leq \theta \leq 1, \lambda \geq -\frac{\beta}{1 - \beta}, \alpha > 0, \beta > 0, 3\alpha^2 (1 + 6\theta^2 + \theta^4) + 2\alpha\beta (1 + \theta^2) + \beta^2 < 1.$$

We assume that κ is taken as the empirical kurtosis and thus by solving for λ we obtain the following functional dependence of this parameter on the other parameters

$$\lambda(\alpha, \beta, \theta) = \frac{1}{1 - m_1} \left(\pm \sigma \sqrt{\frac{3}{(\kappa - 3)(1 - m_2)}} - 1 \right),$$

with the constraint that only α , β and θ such that $\lambda(\alpha, \beta, \theta) \geq -\beta/(1 - \beta)$ are allowed. Then the likelihood as given in (31) is dependent only on \mathbf{a} and $(\alpha_0, \alpha, \beta, \theta)$, since $\delta = 2$ and λ is evaluated from the above equation. The optimization of this function should be performed over the region where the fourth moment of the data exists, i.e. the region described in the Appendix, Propositions 3 and 4.

The original AP-GARCH model is obtained by setting $\lambda = 1$. In this case, one can obtain explicit form for α as function of β and θ that is obtained by solving in for α the following equation

$$\frac{\kappa - 3}{3} = \frac{\sigma^2}{1 - m_2}.$$

After tedious but straightforward calculations, we obtain

$$\alpha(\beta, \theta) = \frac{\sqrt{(\kappa - 1)(1 + \theta^4) + 2(3\kappa - 1)\theta^2 - \frac{2}{3}\kappa\beta^2(1 + \theta^4 + 8\theta^2)} - (\frac{\kappa}{3} - 1)\beta(1 + \theta^2)}{(\kappa - 1)(1 + \theta^4) + 2(3\kappa - 1)\theta^2}.$$

Now, the maximization of the likelihood is with respect of \mathbf{a} and $(\alpha_0, \beta, \theta)$ with the constraints given through $-1 \leq \theta \leq 1$, $0 \leq \beta < 1$ and

$$0 \leq \alpha(\beta, \theta) < \frac{\sqrt{\beta^2(1 + \theta^2)^2 + 3(1 - \beta^2)(1 + 6\theta^2 + \theta^4)} - \beta(1 + \theta^2)}{3(1 + 6\theta^2 + \theta^4)}.$$

6 Empirical data analysis

In this section, we implement different estimation strategies for empirical analysis. As discussed before, the maximum quasi-likelihood method may result in a fit for which empirical characteristics do not match their theoretical counterparts. To ensure this match, one may prefer the maximization of the likelihood with additional constraints. Here, for illustration and also because of its practical significance, we focus on the kurtosis. In our analysis we considered several data sets that are characterized by moderate to fairly large kurtoses. These are: the S&P500 historical data from 3 January 1928 to 30 August 1991 and selected European countries indices: France, Germany, Greece, Italy, Spain, Switzerland, and UK for the period from 2 January 1990 to 31 December 2013. The S&P500 data often serves as a benchmark data with very heavy tail behavior (see also [13] and references therein). The other data sets are milder in terms of heavy tailed behavior but still far from being Gaussian. In the following we discuss effectiveness for both general and constrained maximum quasi-likelihood methods.

Table 2: **Estimated parameters of the APARCH model.**

Data	Estimates					Kurtosis	
	$\hat{\alpha}_0$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\theta}$	$\hat{\delta}$	$\hat{\kappa}_{emp}$	$\hat{\kappa}_{theo}$
S&P500	$10e^{-6}$ ($1e^{-6}$)	.086 (.072)	.914 (.911)	.34 (.29)	1.544 (2)	26.1	12.5 (13.6)
France	$2e^{-5}$ ($3e^{-6}$)	.073 (.063)	.906 (.904)	.54 (.45)	1.584 (2)	9.1	5.6 (6.3)
Germany	$1e^{-5}$ ($4e^{-6}$)	.078 (.073)	.890 (.891)	.41 (.36)	1.815 (2)	11.6	4.9 (5.2)
Greece	$7e^{-6}$ ($6e^{-6}$)	.110 (.107)	.877 (.877)	.08 (.07)	1.982 (2)	7.5	7.5 (12.9)
Italy	$7e^{-6}$ ($4e^{-6}$)	.087 (.084)	.896 (.895)	.26 (.23)	1.875 (2)	7.8	6.0 (7.8)
Spain	$3e^{-5}$ ($4e^{-6}$)	.079 (.069)	.904 (.904)	.38 (.29)	1.563 (2)	8.9	4.7 (4.9)
Switzerland	$2e^{-5}$ ($4e^{-6}$)	.074 (.066)	.890 (.887)	.49 (.43)	1.694 (2)	7.7	4.5 (4.4)
UK	$2e^{-4}$ ($2e^{-6}$)	.060 (.061)	.910 (.910)	.37 (.36)	2.037 (2)	11.9	4.6 (4.7)

Maximum likelihood for the APARCH model

First we consider the estimated parameters of the APARCH model with Gaussian distribution for the noise process obtained through the standard (Gaussian) maximum likelihood method. In Table 2, the parameter estimates are presented. All parameters are found to be significant at 95% level (for brevity the p values are not reported here). A mild level of leverage, as described by parameter $\hat{\theta}$, is noticed and the estimate of the power parameter $\hat{\delta}$ are in all cases closer to 2 than to 1. There is no analytical formula for the theoretical kurtosis for fractional values of δ , thus one has to resort to numerical approximation. Here we use the Monte Carlo method based on 1000 simulated samples from the model and resulted mean values are listed in the last column, while the empirical kurtoses are reported in the second last column. We

Table 3: **Estimated parameters of the extended APARCH model.**

Data	Estimates						Kurtosis	
	$\hat{\alpha}_0$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\theta}$	$\hat{\delta}$	$\hat{\lambda}$	$\hat{\kappa}_{emp}$	$\hat{\kappa}_{theo}$
S&P500	$8e^{-6}$	0.082	0.922	0.361	1.402	0.020	26.12	12.76
France	$4e^{-3}$	0.065	0.925	0.590	1.300	0.017	9.08	5.05
Germany	$7e^{-3}$	0.067	0.926	0.556	1.179	0.015	11.58	4.74
Greece	$6e^{-3}$	0.081	0.925	0.117	1.544	0.002	7.46	8.25
Italy	$1e^{-3}$	0.082	0.913	0.281	1.550	0.018	7.84	5.70
Spain	$3e^{-3}$	0.070	0.925	0.454	1.224	0.029	8.92	4.73
Switzerland	$2e^{-3}$	0.069	0.904	0.549	1.476	0.019	8.92	4.29
UK	$1e^{-3}$	0.058	0.933	0.551	1.396	0.022	11.89	4.85

have seen before that for large values of kurtosis, the empirical kurtosis tends to be very biased and underestimates the true one, so our MC simulation is also affected by this. To provide more accurate approximation of the true kurtosis one would have to correct for bias that is not explored in the present work. It should be also noted that for fractional δ , we do not have explicit conditions for the existence of kurtosis, so, in principle, our estimated parameters may lead to the infinite kurtosis. However, in our simulations, we checked for the stability of the estimates that suggested that the existence condition does not appear to be violated for any of the data set.

For comparison we have also run the estimation by constraining $\delta = 2$. The reason for this choice of δ is the explicit expression for the kurtosis, see the discussion in Example 4. Those estimates and theoretically evaluated kurtoses are reported in the parentheses.

It can be easily observed that none of the data seems to satisfy the assumption of normality – all the empirical kurtosis are beyond the nominal level of 3. Moreover, it is worth noticing that the kurtosis ($\hat{\kappa}_{theo}$) captured by the estimated parameters is far from the empirical kurtosis. The same applies to the model with δ fixed to 2, which also shows very similar fit of the parameters. This indicates again that the parameter δ is not that influential. At the same time, it shows the fit parameters do not fully translate into the heavy-tailedness exhibited in the data.

Maximum likelihood for the extended APARCH model

In another study, we estimate our extended model defined in (29) and investigate the effect of introducing a scale parameter λ . The results in Table 3 summarize the parameter estimates of the extended APARCH model. There are few interesting points to be made. Firstly, the estimated values of $\hat{\lambda}$ are close of 0 so the assumed value of 1 in the original model is not supported by the data. Secondly, by introducing λ a drop in the values of $\hat{\delta}$ is noticed and this drop is quite significant for some of the series.

Table 4: **Estimates for the extended APARCH model with $\delta = 1$ ($\delta = 2$)**

Data	Estimates					Kurtosis	
	$\hat{\alpha}_0$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\theta}$	$\hat{\lambda}$	$\hat{\kappa}_{emp}$	$\hat{\kappa}_{theo}$
S&P500	$6e^{-3}$ ($4e^{-5}$)	.085 (.070)	.924 (.914)	.42 (.29)	.019 (.027)	26.1	19.6 (15.4)
France	$1e^{-2}$ ($2e^{-4}$)	.066 (.056)	.930 (.916)	.64 (.45)	.018 (.013)	9.1	5.0 (6.3)
Germany	$1e^{-2}$ ($2e^{-4}$)	.068 (.057)	.928 (.916)	.59 (.40)	.016 (.011)	11.6	4.9 (5.2)
Greece	$4e^{-2}$ ($2e^{-3}$)	.073 (.072)	.936 (.917)	.14 (.11)	.003 (.002)	7.7	8.6 (6.9)
Italy	$1e^{-2}$ ($2e^{-4}$)	.078 (.077)	.922 (.906)	.37 (.23)	.018 (.018)	7.8	4.9 (8.9)
Spain	$9e^{-3}$ ($9e^{-5}$)	.070 (.064)	.927 (.913)	.51 (.31)	.028 (.034)	8.9	4.8 (5.8)
Switzerland	$1e^{-2}$ ($2e^{-4}$)	.072 (.061)	.912 (.894)	.61 (.47)	.022 (.016)	7.7	4.2 (4.6)
UK	$8e^{-3}$ ($7e^{-5}$)	.059 (.051)	.939 (.926)	.66 (.41)	.012 (.023)	11.9	5.1 (5.4)

Thirdly, the theoretical kurtosis are evaluated through the estimated parameters.

In the next analysis, we proceed by fixing the parameter δ to be either 1 or 2 and estimating the rest of the parameters of the extended model. Table 4 summarizes the results with estimated values for the case of $\delta = 2$ reported in parentheses. It shows that the parameter δ is not contributing much to the tail when parameters are estimated through the standard maximum likelihood method. This is in contrast to what was argued in [5] and shows that in the extended model the role of δ is, to some extent, played by λ . In Tables 3 and 4, however, the estimated model parameters lack the capability of accounting for the kurtosis observed in data, in all except one cases the later being bigger than the ones evaluated through the theoretical formula (or the MC method).

In the nutshell, we conclude that using estimated δ makes analytical formulas for kurtosis unavailable. Moreover, as seen in Figure 10, evaluating the kurtosis by using the Monte Carlo method is difficult due to its bias and large variance. On the other hand if instead the actual estimate one uses $\delta = 1$ or $\delta = 2$, then both the original and extended models perform well and have explicit analytical formulas available. To capture the sample kurtosis even better one can resort to the constrained method that is applied to the data next.

Constrained maximum likelihood estimation

To illustrate how to use the constrained method to account for the heavy tails in the considered data sets, we restrict ourselves to $\lambda = 1$, $\delta = 2$ and $\theta = 0$. This choice brings us to simplicity of the formulas derived in the last part of Section 5.2.

Table 5 summarizes the results obtained through two different estimation strategies (see Section 5). The first three columns of the table report the estimates of parameters obtained through the standard likelihood method. The next three columns present the estimates via the proposed hybrid method. It can be seen that the parameter

Table 5: Estimation of the APARCH model through constrained likelihood

Data	Standard Likelihood Estimates			Constrained Likelihood Estimates			Kurtosis		
	$\hat{\alpha}_0$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\beta}$	$\hat{\kappa}_{emp}$	$\hat{\kappa}_{theo}$	$\hat{\kappa}_{c,theo}$
S&P500	$1e^{-6}$	0.0832	0.9061	$10e^{-7}$	0.0841	0.9079	26.12	8.58	26.70
<i>France</i>	$3e^{-6}$	0.0870	0.8991	$3e^{-6}$	0.0979	0.8877	9.08	6.64	9.10
<i>Germany</i>	$3e^{-6}$	0.0906	0.8913	$3e^{-6}$	0.0983	0.8886	11.58	5.53	11.64
<i>Greece</i>	$6e^{-6}$	0.1049	0.8816	$6e^{-6}$	0.0974	0.8866	7.46	16.92	7.46
<i>Italy</i>	$3e^{-6}$	0.0929	0.8952	$4e^{-6}$	0.0970	0.8877	7.84	11.09	7.88
<i>Spain</i>	$3e^{-6}$	0.0835	0.9025	$3e^{-6}$	0.0867	0.9019	8.92	6.02	8.90
<i>Switzerland</i>	$3e^{-6}$	0.0858	0.8901	$4e^{-6}$	0.1198	0.8565	7.71	4.34	7.74
<i>UK</i>	$2e^{-6}$	0.0815	0.9030	$2e^{-6}$	0.0900	0.8991	11.89	5.28	11.86

estimates obtained by these two methods are very close to each other. However, their effect on accounting for the observed kurtosis is significantly different. The last part of the table shows the empirical kurtosis ($\hat{\kappa}_{emp}$), the theoretical kurtosis obtained via the estimates from standard log-likelihood method ($\hat{\kappa}_{theo}$) and the theoretical kurtosis obtained via the estimates from the constrained log-likelihood method ($\hat{\kappa}_{c,theo}$). The results clearly show the improvement that has been achieved in accounting for the true tail behaviour of data. Explanation for the parameter estimates of not being significantly apart for the two methods, while the kurtoses being noticeably different, lies in the sensitivity of the tail heaviness near the boundary of the model existence region.

7 Conclusions

Our detailed study of the extended APARCH model allows us to quantify the effect of the model parameters on the dependence and the distributional tails. In particular, we obtain conditions for the existence of the model as well as an explicit formula for the correlation, moments, and leverage effect. From these explicit conditions, one can determine the range of the parameters for which particular characteristics, such as kurtosis, leverage correlation or autocorrelation, are formally meaningful. By using the derived results, we have been able, firstly, to analyze importance of the parameters and, secondly, to implement the constrained likelihood method of data fitting that allows us to preserve accurately the selected ‘stylized facts’.

Additionally, we have assessed the role of the power parameter δ . We show that the obtained formulas are particularly useful for the case of $\delta = 1$ and $\delta = 2$, when they allow for evaluation of important theoretical characteristics of the model. For other values of δ , they yield explicit results only for the δ powers of returns and volatility.

Consequently, the effect of δ parameter is difficult to assess because the effect seen for the δ powers may differ significantly from that for the actual returns. For example, the presented Monte Carlo studies show that the dependence in the actual returns as measured by the correlation coefficient can be at the level 0.6 while for the δ power the correlation is nearly one, when δ approaches the value of two.

In the literature, it was argued that the power parameter seem to play a similar role in controlling the tails of the distribution as the power in the Box-Cox transformation. We did not found any conclusive evidence of this and the effect on the tails seem to be rather moderate. However, the parameter δ does seem to play a dual role as it affects both the kurtosis and autocorrelation of ρ_t . Based our findings, we recommend limiting the model to two natural values of δ : one and two, and using our extended model with the additional parameter λ to control the time dependence in the volatility. For practical reasons, this model seems to be equally flexible in accounting for the features in the data, while analytically and numerically it is more trackable and easier to study. This was also demonstrated in our analysis of empirical data by showing that both the parameters and the kurtoses are not greatly affected by simply setting δ to the fixed value of either one or two.

We considered two conditionally Gaussian likelihood-based methods of evaluation. The first one is the standard maximum (conditional) likelihood. The second method additionally takes into account empirical characteristics of the data such as sample kurtosis and leverage to constrain the parameter space over which the likelihood is maximized. Both the methods seem to perform well for data simulated from the model but the constrained method may prove beneficial in practice. More detailed statistical analysis of the method and of effects of the choice of constraints is needed.

The obtained results on the APARCH model provide concrete technical tools both for further theoretical studies and for utilization of the model to investigate financial time series. Our studies demonstrate the importance of careful theoretical investigation of the model in order to draw accurate conclusions about its ability to account for features and the so-called ‘stylized facts’ observed in the real data.

Appendix

Recurrent formula for moments of λ_t

To evaluate moments of the model, we use the integer moments of λ_t defined by

$$\lambda_t = \alpha [(1 - \theta)^\delta e_t^{+\delta} + (1 + \theta)^\delta e_t^{-\delta}] + \beta. \quad (35)$$

Our formula involves $e(p) = \text{E}e_t^{+p} = \text{E}e_t^{-p}$ given in the following

$$e(p) = \frac{2^{(p-1)/2}}{\sqrt{2} \pi} \int_0^\infty x^{(p-1)/2} e^{-x} dx = \frac{2^{p/2-1}}{\sqrt{\pi}} \Gamma\left(\frac{p+1}{2}\right). \quad (36)$$

We formulate this as a lemma and skip the proof which is straightforward.

Lemma 1. *The integer valued moments $m_k = E(\lambda_t^k)$ of λ_t can be obtained from the following recurrent relation*

$$m_k = \sum_{r=0}^k \binom{k}{r} \alpha^r ((1-\theta)^{r\delta} + (1+\theta)^{r\delta}) e(r\delta) \beta^{k-r}.$$

Corollary 1. *Let m_1 and m_2 be the first and second moment of λ_t . Then*

$$\begin{aligned} m_1 &= \Gamma\left(\frac{\delta+1}{2}\right) \frac{\alpha [(1-\theta)^\delta + (1+\theta)^\delta] 2^{\delta/2-1}}{\sqrt{\pi}} + \beta, \\ m_2 &= \frac{2^{\delta-1} \alpha^2}{\sqrt{\pi}} ((1-\theta)^{2\delta} + (1+\theta)^{2\delta}) \Gamma\left(\delta + \frac{1}{2}\right) + \\ &\quad + \frac{2^{\delta/2} \alpha \beta}{\sqrt{\pi}} ((1-\theta)^\delta + (1+\theta)^\delta) \Gamma\left(\frac{\delta+1}{2}\right) + \beta^2. \end{aligned}$$

Moreover, variance σ^2 of λ_t is given as

$$\sigma^2 = \frac{2^{\delta-1} \alpha^2}{\pi} \Gamma^2\left(\frac{\delta+1}{2}\right) \left(\left(\sqrt{\pi} \frac{\Gamma(\delta + \frac{1}{2})}{\Gamma^2(\frac{\delta+1}{2})} - \frac{1}{2} \right) ((1-\theta)^{2\delta} + (1+\theta)^{2\delta}) - (1-\theta^2)^\delta \right)$$

Stationarity condition for the volatility model

We provide with the conditions for the existence of a stationary solution to the following series, which also satisfies the recurrence relation (3):

$$\rho_t^\delta = \alpha_0 \left(\lambda + \sum_{k=1}^{\infty} \lambda_{t-1} \cdots \lambda_{t-k} \right). \quad (37)$$

We consider convergence of the above series in the mean-square sense and, more generally, in L_q -sense, where $q > 0$ and L_q is the space of random variables with the finite q moment. It is well known that the L_q -spaces are Banach spaces and thus the absolute convergence of the series implies its convergence in the L_q -norm. We note that by proving the L_q convergence for ρ_t^δ we show that ρ_t belongs to L_p space with $p = q\delta$. The absolute convergence means that

$$\sum_{k=1}^{\infty} (E(\lambda_{t-1} \cdots \lambda_{t-k})^q)^{1/q} < \infty$$

and since $E(\lambda_{t-1} \cdots \lambda_{t-k})^q = m_q^k$, where $m_q = E\lambda_t^q$, the condition for convergence reduces to $m_q < 1$. By Hölder's inequality, we note that if $0 < p < q$, then $m_p < 1$ is less restrictive than $m_q < 1$.

Note that by the triangle inequality for the q -norm

$$(E(\lambda_t^q))^{\frac{1}{q}} \leq \alpha ((1-\theta)^\delta + (1+\theta)^\delta) e^{\frac{1}{q}}(p) + 2\beta.$$

From this observation we obtain immediately the following result.

Proposition 2. *A sufficient condition for existence of the strictly stationary solution ρ_t in (1) such that it belongs to L_p is given by the inequality*

$$\alpha \left((1 - \theta)^\delta + (1 + \theta)^\delta \right) e^{\frac{1}{q}}(p) + 2\beta < 1, \quad (38)$$

where $q = p/\delta$ and α, β, δ and θ are parameters in the model given in (1), while $m(p)$ is given in (36). We note that for a symmetric case ($\theta = 0$) we obtain simply

$$\alpha e^{\frac{1}{q}}(p) + \beta < \frac{1}{2}. \quad (39)$$

For the important special case of $q = 1$, we have the explicit value for $E(\lambda_t)^q$ given in Corollary 1. This allows to obtain the sufficient and necessary condition for the absolute convergence of the series defining ρ_t^δ in the L_p -norm. Namely, we have the following result.

Proposition 3. *A sufficient and necessary condition for the existence of a strictly stationary solution ρ_t to (1) that belongs to L_δ (in the absolute convergence of the series) is given by the inequality for the parameters $\alpha > 0$, $\beta \in [0, 1]$, and $\delta > 0$:*

$$(1 - \theta)^\delta + (1 + \theta)^\delta < \sqrt{\pi} \frac{1 - \beta}{\alpha} \frac{2^{1-\delta/2}}{\Gamma\left(\frac{\delta+1}{2}\right)}. \quad (40)$$

We also note two important special cases. Firstly, $\delta = 1$ yields the condition that does not depend on θ :

$$1 < \sqrt{\frac{\pi}{2}} \frac{1 - \beta}{\alpha}. \quad (41)$$

Secondly, the case of $\delta = 2$ yields

$$1 + \theta^2 < \frac{1 - \beta}{\alpha}. \quad (42)$$

Similarly, we can obtain a stronger sufficient and necessary condition for the case $q = 2$, for which we have also exact value for the moment in Corollary 1.

Proposition 4. *A sufficient and necessary condition for the existence of a strictly stationary solution ρ_t to (1) that belongs to $L_{2\delta}$ (in the absolute convergence of the series) is given by the inequality*

$$\begin{aligned} 2^{\delta/2-1} \Gamma\left(\delta + \frac{1}{2}\right) \alpha \left((1 - \theta)^{2\delta} + (1 + \theta)^{2\delta} \right) + \Gamma\left(\frac{\delta+1}{2}\right) \beta \left((1 - \theta)^\delta + (1 + \theta)^\delta \right) < \\ < \sqrt{\frac{\pi}{2^\delta}} \frac{1 - \beta^2}{\alpha}. \end{aligned} \quad (43)$$

We note two important special cases. Firstly, $\delta = 1$ yields the condition

$$\alpha(1 + \theta^2) + \frac{2^{3/2}}{\sqrt{\pi}} \beta < \frac{1 - \beta^2}{\alpha} \quad (44)$$

Secondly, the case of $\delta = 2$ yields

$$3\alpha(1 + 6\theta^2 + \theta^4) + 2\beta(1 + \theta^2) < \frac{1 - \beta^2}{\alpha}. \quad (45)$$

In our estimation strategies, we have emphasized the importance to limit the range of parameters so appropriate moments (the kurtosis, for example) exist. It is clear that the n th moment of y_t exists whenever the n th moment of ρ_t is finite. The fourth moment is needed to capture tails through kurtosis. For $\delta = 1$, we can use Proposition 2 with $p = 4$ that leads to

$$2\alpha^4 \sqrt{\frac{3}{2}} + 2\beta < 1. \quad (46)$$

However, in order to have flexibility in modeling kurtosis and thus tails of the data, we provide a stronger result for this case that is based on exact formula for the fourth moment given in (18) to obtain a lengthy but elementary sufficient condition.

Proposition 5. *For the case $\delta = 1$, volatility ρ_t exists in the L_4 sense if*

$$\beta^4 + 4\alpha\beta^3 \sqrt{\frac{2}{\pi}} + 6\alpha^2\beta^2(1 + \theta^2) + 8\alpha^3\beta(1 + 3\theta^2) \sqrt{\frac{2}{\pi}} + 3\alpha^4(1 + 6\theta^2 + \theta^4) \leq 1.$$

On the other hand, for $\delta = 2$ one can use the sufficient condition for finite kurtosis given in (45) listed in Proposition 4.

For the leverage effect, restrictions on the parameters depend on what measure of the leverage is considered. This was discussed in Subsection 4.2. There for $\delta = 1$, two convenient characteristics were explicitly evaluated: $r(\rho_t, \epsilon_{t-1})$ and $r(\rho_t^2, \epsilon_{t-1})$. They require only the second and third moment of ρ_t , respectively. Thus the condition following from the existence of kurtosis implies existence of both leverage characteristics. For $\delta = 2$, the convenient characteristics are $r(\rho_t^2, r_{t-1}^2)$ and $r(\rho_t^4, r_{t-1}^2)$. For the first one the existence of the kurtosis suffices, while for the second the sixth moment is needed. For the latter one can write the parameter restriction using Proposition 2.

Moments of ρ_t

Here we collect some results on the moments of ρ_t that are used throughout the paper. First, by the power of a sum algebraic formula we have a relation between these moments and moments of the series $L = \sum_{k=1}^{\infty} \lambda_{-1} \dots \lambda_t$.

Proposition 6. *We have*

$$E(\rho_t^{k\delta}) = \alpha_0^k \sum_{r=0}^k \binom{k}{r} \lambda^r M_{k-r},$$

where M_k is the k -th moment of the series L .

For computing moments explicitly in terms of the actual parameters of the model one can utilize the following lemma together with Lemma 1.

Lemma 2. *Let m_k be the k -th moment of λ_t . Then*

$$\begin{aligned} M_1 &= p_1, \\ M_2 &= p_2(1 + 2p_1), \\ M_3 &= p_3 \left(1 + 3(p_1 + p_2) + 6p_1p_2 \right), \\ M_4 &= p_4 \left(1 + 4(p_1 + p_3) + 6p_2 + 12(p_1p_2 + p_1p_3 + p_2p_3) + 24p_1p_2p_3 \right), \end{aligned}$$

where $p_k = m_k/(1 - m_k)$.

Proof. The proof of the first and second moments can be seen from the previous results. For the sake of brevity, we just present the fourth moment argument – the third moment can be obtained in a similar fashion. We note that after some combinatorics and algebra, the fourth moment can be simplified as

$$L^4 = \sum_{k,j,l,m=1}^{\infty} (\lambda_{-1} \dots \lambda_{-k}) (\lambda_{-1} \dots \lambda_{-j}) (\lambda_{-1} \dots \lambda_{-l}) (\lambda_{-1} \dots \lambda_{-m}) = \sum_{i=1}^{\infty} (\lambda_{-1} \dots \lambda_{-i})^4 L_i,$$

where

$$\begin{aligned} L_i &= 1 + 4 \sum_{m=1}^{\infty} \lambda_{-i-1}^3 \dots \lambda_{-i-m}^3 + 4 \sum_{m=1}^{\infty} \lambda_{-i-1} \dots \lambda_{-i-m} + 6 \sum_{m=1}^{\infty} \lambda_{-i-1}^2 \dots \lambda_{-i-m}^2 \\ &\quad + 12 \left(\sum_{m=1}^{\infty} \lambda_{-i-1}^2 \dots \lambda_{-i-m}^2 \sum_{n=1}^{\infty} \lambda_{-i-m-1} \dots \lambda_{-i-m-n} \right. \\ &\quad + \sum_{m=1}^{\infty} \lambda_{-i-1}^3 \dots \lambda_{-i-m}^3 \sum_{n=1}^{\infty} \lambda_{-i-m-1} \dots \lambda_{-i-m-n} \\ &\quad \left. + \sum_{m=1}^{\infty} \lambda_{-i-1}^3 \dots \lambda_{-i-m}^3 \sum_{n=1}^{\infty} \lambda_{-i-m-1}^2 \dots \lambda_{-i-m-n}^2 \right) \\ &\quad + 24 \sum_{m=1}^{\infty} \lambda_{-i-1}^3 \dots \lambda_{-i-m}^3 \sum_{n=1}^{\infty} \lambda_{-i-m-1}^2 \dots \lambda_{-i-m-n}^2 \sum_{l=1}^{\infty} \lambda_{-i-m-n-1} \dots \lambda_{-i-m-n-l} \end{aligned}$$

Applying the expectations we get,

$$\begin{aligned}
L^4 &= \sum_{i=1}^{\infty} m_4^i \left(1 + 4 \sum_{m=1}^{\infty} m_3^m + 4 \sum_{m=1}^{\infty} m_1^m + 6 \sum_{m=1}^{\infty} m_2^m + 12 \left(\sum_{m=1}^{\infty} m_2^m \sum_{n=1}^{\infty} m_1^n \right. \right. \\
&\quad \left. \left. + \sum_{m=1}^{\infty} m_3^m \sum_{n=1}^{\infty} m_1^n + \sum_{m=1}^{\infty} m_3^m \sum_{n=1}^{\infty} m_2^n \right) + 24 \left(\sum_{m=1}^{\infty} m_3^m \sum_{n=1}^{\infty} m_2^n \sum_{l=1}^{\infty} m_1^l \right) \right) \\
&= \frac{m_4}{1-m_4} \left[1 + 4 \left(\frac{m_3}{1-m_3} + \frac{m_1}{1-m_1} \right) + 6 \frac{m_2}{1-m_2} + 12 \left(\frac{m_1}{1-m_1} \frac{m_2}{1-m_2} \right. \right. \\
&\quad \left. \left. + \frac{m_1}{1-m_1} \frac{m_3}{1-m_3} + \frac{m_2}{1-m_2} \frac{m_3}{1-m_3} \right) + 24 \frac{m_1}{1-m_1} \frac{m_2}{1-m_2} \frac{m_3}{1-m_3} \right].
\end{aligned}$$

□

Autocorrelation of volatility and heteroskedastic innovations

It is important for both theoretical and practical considerations to have insight into time dependence in our volatility model ρ_t as well as in the conditionally heteroskedastic (CH) innovations $\epsilon_t = \rho_t e_t$. The explicit formulas for the correlations of these two process are not trackable for arbitrary δ . However one can get relatively simple formulas for autocorrelations of ρ_t^δ and ϵ_t^δ . Some results on the autocorrelation of ϵ_t for the Gaussian innovation can be found in [11]. Our derivation here are more general since they cover our extended scale-location model (2), including non-Gaussian innovations. Moreover, we give the autocorrelation of the powers of volatility ρ_t .

Proposition 7. *Let random variables λ_i 's in (2) be iid such that $\sigma^2 + m^2 < 1$, where m and σ^2 are their mean and variance, respectively. Then for $t \geq 0$ the correlation of ρ_t^δ is given by $r(\rho_t^\delta, \rho_0^\delta) = m^t$.*

Proof. It is enough to consider the case $\alpha_0 = 1$ for which we have

$$\begin{aligned}
\text{Cov}(\rho_t^\delta, \rho_0^\delta) &= \text{Cov} \left(\sum_{k=1}^{\infty} \lambda_{t-1} \cdots \lambda_{t-k}, \sum_{k=1}^{\infty} \lambda_{-1} \cdots \lambda_{-k} \right) \\
&= \text{Cov} \left(\sum_{k=1}^t \lambda_{t-1} \cdots \lambda_{t-k} + \sum_{k=t+1}^{\infty} \lambda_{t-1} \cdots \lambda_{t-k}, \sum_{k=1}^{\infty} \lambda_{-1} \cdots \lambda_{-k} \right) \\
&= \text{Cov} \left(\sum_{k=1}^t \lambda_{t-1} \cdots \lambda_{t-k}, \sum_{k=1}^{\infty} \lambda_{-1} \cdots \lambda_{-k} \right) + \\
&\quad + \text{Cov} \left(\sum_{k=t+1}^{\infty} \lambda_{t-1} \cdots \lambda_{t-k}, \sum_{k=1}^{\infty} \lambda_{-1} \cdots \lambda_{-k} \right)
\end{aligned}$$

$$\begin{aligned}
&= \mathbb{E}(\lambda_{t-1} \cdots \lambda_0) \text{Var} \left(\sum_{k=1}^{\infty} \lambda_{-1} \cdots \lambda_{-k} \right) \\
&= m^t \text{Var}(\rho_t^\delta),
\end{aligned}$$

where the last two equations are due to independence between λ_i 's. \square

We can also give the correlation structure for the δ power of the absolute value of the innovations $\epsilon_t = \rho_t e_t$ in APARCH models with unspecified noise e_t .

Proposition 8. *Let random variables λ_i 's be as in (2). By m and σ^2 we denote their mean and variance, respectively. Additionally we assume that λ_t is a non-random function of e_t , where e_t are iid random variables. Denote $\gamma_\delta = \text{Cov}(\lambda_0, |e_0|^\delta)$ and $\nu_\delta = \mathbb{E}(|e_0|^\delta)$. Then the autocorrelation of $|\epsilon_t|^\delta = \rho_t^\delta |e_t|^\delta$ is given*

$$r(|\epsilon_t|^\delta, |\epsilon_0|^\delta) = \frac{(1 - m^2) \gamma_\delta + \sigma^2 m \nu_\delta}{\sigma^2 \nu_{2\delta} + (\lambda + m(1 - \lambda))^2 (1 - \sigma^2 - m^2) (\nu_{2\delta} - \nu_\delta^2)} \nu_\delta m^{t-1},$$

and

$$\text{Var}(|\epsilon_t|^\delta) = \left(\frac{\alpha_0}{1 - m} \right)^2 \frac{\sigma^2 \nu_{2\delta} + (\lambda + m(1 - \lambda))^2 (1 - \sigma^2 - m^2) (\nu_{2\delta} - \nu_\delta^2)}{1 - \sigma^2 - m^2}.$$

Proof. Let us start with computing variance of $|\epsilon_t|^\delta$. By independence of ρ_0 and e_0 we have

$$\begin{aligned}
\text{Var}(\rho_0^\delta |e_0|^\delta) &= \mathbb{E} \rho_0^{2\delta} \mathbb{E} |e_0|^{2\delta} - (\mathbb{E} \rho_0^\delta)^2 (\mathbb{E} |e_0|^\delta)^2 \\
&= \text{Var}(\rho_0^\delta) \nu_{2\delta} + (\nu_{2\delta} - \nu_\delta^2) (\mathbb{E} \rho_0^\delta)^2 \\
&= \left(\frac{\alpha_0}{1 - m} \right)^2 \frac{\sigma^2 \nu_{2\delta} + (\lambda + m(1 - \lambda))^2 (1 - \sigma^2 - m^2) (\nu_{2\delta} - \nu_\delta^2)}{1 - \sigma^2 - m^2}.
\end{aligned}$$

To compute autocovariance note that for $t \geq 1$

$$\begin{aligned}
\text{Cov}(|\epsilon_t|^\delta, |\epsilon_0|^\delta) &= \text{Cov}(\rho_t^\delta |e_t|^\delta, \rho_0^\delta |e_0|^\delta) \\
&= \mathbb{E} \left(|e_t|^\delta \right) \text{Cov}(\rho_t^\delta, \rho_0^\delta |e_0|^\delta) \\
&= \nu_\delta \left(\text{Cov}(\rho_t^\delta, \rho_0^\delta \tilde{e}_\delta) + \nu_\delta \text{Cov}(\rho_t^\delta, \rho_0^\delta) \right),
\end{aligned}$$

where $\tilde{e}_\delta = |e_0|^\delta - \nu_\delta$. From Proposition 7:

$$\begin{aligned}
\text{Cov}(\rho_t^\delta, \rho_0^\delta) &= \text{Var}(\rho_0^\delta) m^t \\
&= \left(\frac{\alpha_0}{1 - m} \right)^2 \frac{\sigma^2}{1 - \sigma^2 - m^2} m^t
\end{aligned}$$

Note that for $t \geq 1$ we can write

$$\sum_{k=1}^{\infty} \lambda_{t-1} \cdots \lambda_{t-k} = \sum_{k=1}^t \lambda_{t-1} \cdots \lambda_{t-k} + \lambda_{t-1} \cdots \lambda_0 \left(1 + \sum_{j=1}^{\infty} \lambda_{-1} \cdots \lambda_{-j} \right),$$

with the first term be independent of both \tilde{e}_δ and ρ_0 . Thus we have for $t \geq 1$:

$$\begin{aligned} \text{Cov}(\rho_t^\delta, \rho_0^\delta \tilde{e}_\delta) &= \alpha_0^2 \text{Cov} \left(\sum_{k=1}^{\infty} \lambda_{t-1} \cdots \lambda_{t-k}, \left(1 + \sum_{k=1}^{\infty} \lambda_{-1} \cdots \lambda_{-k} \right) \tilde{e}_\delta \right) \\ &= \alpha_0^2 \text{Cov} \left(\lambda_{t-1} \cdots \lambda_0 \left(1 + \sum_{k=1}^{\infty} \lambda_{-1} \cdots \lambda_{-k} \right), \sum_{k=1}^{\infty} \lambda_{-1} \cdots \lambda_{-k} \tilde{e}_\delta \right) \\ &\quad + \alpha_0^2 \text{Cov} \left(\lambda_{t-1} \cdots \lambda_0 \left(1 + \sum_{k=1}^{\infty} \lambda_{-1} \cdots \lambda_{-k} \right), \tilde{e}_\delta \right) \\ &= \alpha_0^2 m^{t-1} \text{Cov} \left(\lambda_0 \left(1 + \sum_{k=1}^{\infty} \lambda_{-1} \cdots \lambda_{-k} \right), \sum_{k=1}^{\infty} \lambda_{-1} \cdots \lambda_{-k} \tilde{e}_\delta \right) \\ &\quad + \alpha_0^2 m^{t-1} \text{Cov} \left(\lambda_0 \left(1 + \sum_{k=1}^{\infty} \lambda_{-1} \cdots \lambda_{-k} \right), \tilde{e}_\delta \right) \\ &= \alpha_0^2 m^{t-1} \gamma_\delta \left(\text{E} \left(\sum_{k=1}^{\infty} \lambda_{-1} \cdots \lambda_{-k} \right)^2 + \text{E} \left(\sum_{k=1}^{\infty} \lambda_{-1} \cdots \lambda_{-k} \right) \right) \\ &\quad + \alpha_0^2 m^{t-1} \left(\text{Cov}(\lambda_0, \tilde{e}_\delta) + \text{Cov} \left(\lambda_0 \sum_{k=1}^{\infty} \lambda_{-1} \cdots \lambda_{-k}, \tilde{e}_\delta \right) \right) \\ &= \alpha_0^2 m^{t-1} \gamma_\delta \left(\text{Var} \left(\sum_{k=1}^{\infty} \lambda_{-1} \cdots \lambda_{-k} \right) + \frac{m}{1-m} \left(1 + \frac{m}{1-m} \right) \right) + \\ &\quad \quad \quad + \frac{\alpha_0^2 m^{t-1} \gamma_\delta}{1-m} \\ &= \left(\frac{\alpha_0}{1-m} \right)^2 m^{t-1} \gamma_\delta \left(\frac{\sigma^2}{1-\sigma^2-m^2} + m \right) \\ &= \frac{\alpha_0^2}{(1-m)^2} \frac{(1-m^2)}{1-\sigma^2-m^2} \gamma_\delta m^{t-1}. \end{aligned}$$

Thus combining the two results we obtain the formula for the autocovariance function

$$\text{Cov}(|\epsilon_t|^\delta, |\epsilon_0|^\delta) = \left(\frac{\alpha_0}{1-m} \right)^2 \frac{(1-m^2) \gamma_\delta + \sigma^2 m \nu_\delta}{1-\sigma^2-m^2} \nu_\delta m^{t-1}.$$

□

Corollary 2. *Let us assume that e_0 has a symmetric distribution around zero. Then*

$$\gamma_\delta = \frac{\alpha}{2} \left((1 + \theta)^\delta + (1 - \theta)^\delta \right) (\nu_{2\delta} - \nu_\delta^2)$$

and

$$m = \frac{\alpha}{2} \left((1 + \theta)^\delta + (1 - \theta)^\delta \right) \nu_\delta + \beta,$$

$$\sigma^2 = \frac{\alpha^2}{4} \left(2\nu_{2\delta} \left((1 + \theta)^{2\delta} + (1 - \theta)^{2\delta} \right) - \nu_\delta^2 \left((1 + \theta)^\delta + (1 - \theta)^\delta \right)^2 \right).$$

Proof. Be the assumed symmetry of e_0 , we have

$$\begin{aligned} \text{Cov}(\lambda_0, |e_0|^\delta) &= \alpha \left((1 + \theta)^\delta + (1 - \theta)^\delta \right) \text{Cov}(e_0^{+\delta}, |e_0|^\delta) \\ &= \frac{\alpha}{2} \left((1 + \theta)^\delta + (1 - \theta)^\delta \right) \text{Var}(|e_0|^\delta) \\ &= \frac{\alpha}{2} \left((1 + \theta)^\delta + (1 - \theta)^\delta \right) (\nu_{2\delta} - \nu_\delta^2). \end{aligned}$$

□

Let ρ_t be the time varying volatility process, y_t be the returns in the APARCH model (1) with a constant function f and the δ -powers returns be defined as $r_t^\delta = y_t^{+\delta} - y_t^{-\delta}$. Then the correlation between the δ -powers of volatility, ρ_t^δ , and lagged centered returns, $\epsilon_{t-1}^{(\delta)}$, can be regarded as a measure of the leverage effect. Alternatively, one can view the square of standard deviation (variance) as the measure of volatility, so that the leverage effect can be described through the correlation between of $\rho_t^{2\delta}$ and the lagged returns. Denote as before $\gamma_\delta = \text{Cov}(\lambda_0, |e_0|^\delta)$. The explicit relations for these correlation are given in the following result.

Proposition 9. *Let us assume that the random variables e_t have a distribution that is symmetric around zero. With the notation introduced above we have*

$$r(\rho_t^\delta, \epsilon_{t-1}^{(\delta)}) = \alpha \sqrt{\nu_{2\delta}(1 + \text{cv}_1^{-2})} \frac{(1 - \theta)^\delta - (1 + \theta)^\delta}{2},$$

$$r(\rho_t^{2\delta}, \epsilon_{t-1}^{(\delta)}) = 2r(\rho_t^\delta, \epsilon_{t-1}^{(\delta)}) \frac{\text{cv}_1}{\text{cv}_2} \left(\alpha_0 + \left(\alpha \frac{(1 - \theta)^\delta + (1 + \theta)^\delta}{2} \frac{\nu_{3\delta}}{\nu_{2\delta}} + \beta \right) \frac{\text{E}(\rho_0^{3\delta})}{\text{E}(\rho_0^{2\delta})} \right) \frac{\text{E}(\rho_0^\delta)}{\text{E}(\rho_0^{2\delta})},$$

where cv_1 is the coefficient of variation for ρ_0^δ while cv_2 is the coefficient of variation of $\rho_0^{2\delta}$.

Proof. We note that because the expected value of ϵ_{t-1} is zero (the symmetry of e_0):

$$\begin{aligned} \text{Cov}(\rho_t^{k\delta}, \epsilon_{t-1}^{(\delta)}) &= \text{Cov} \left((\alpha_0 + \rho_0^\delta \lambda_0)^k, \rho_0^\delta (e_0^{+\delta} - e_0^{-\delta}) \right) \\ &= \text{E} \left(\rho_0^\delta \text{Cov} \left((\alpha_0 + \rho_0^\delta \lambda_0)^k, e_0^{+\delta} - e_0^{-\delta} \right) \right) \end{aligned}$$

For $k = 1$ and through conditioning on ρ_0 that is independent of e_0 and of λ_0 , we have

$$\begin{aligned} \mathbb{E} \left(\rho_0^\delta \text{Cov} \left(\alpha_0 + \rho_0^\delta \lambda_0, e_0^{+\delta} - e_0^{-\delta} \right) \right) &= \alpha \mathbb{E} \left(\rho_0^{2\delta} \right) \text{Cov} \left((1 - \theta)^\delta e_0^{+\delta} + (1 + \theta)^\delta e_0^{-\delta}, e_0^{+\delta} - e_0^{-\delta} \right) \\ &= \alpha \mathbb{E} \left(\rho_0^{2\delta} \right) \mathbb{E} \left((1 - \theta)^\delta e_0^{+2\delta} - (1 + \theta)^\delta e_0^{-2\delta} \right) \\ &= \frac{\alpha}{2} \mathbb{E} \left(\rho_0^{2\delta} \right) \left((1 - \theta)^\delta - (1 + \theta)^\delta \right) \nu_{2\delta}. \end{aligned}$$

The first part of the result follows if we note that

$$\begin{aligned} \text{Var}(\rho_0^\delta) &= \mathbb{E}(\rho_0^{2\delta}) \frac{\text{cv}_1^2}{1 + \text{cv}_1^2}, \\ \text{Var}(r_0^\delta) &= \mathbb{E}(\rho_0^{2\delta}) \nu_{2\delta}. \end{aligned}$$

For $k = 2$, let us note that

$$\begin{aligned} \mathbb{E} \left(\rho_0^\delta \text{Cov} \left((\alpha_0 + \rho_0^\delta \lambda_0)^2, e_0^{+\delta} - e_0^{-\delta} \right) \right) &= \alpha_0 \alpha \mathbb{E} \left(\rho_0^{2\delta} \right) \left((1 - \theta)^\delta - (1 + \theta)^\delta \right) \nu_{2\delta} + \\ &\quad + \mathbb{E} \left(\rho_0^{3\delta} \right) \mathbb{E} \left(\left(\alpha \left((1 - \theta)^\delta e_0^{+\delta} + (1 + \theta)^\delta e_0^{-\delta} \right) + \beta \right)^2 \left(e_0^{+\delta} - e_0^{-\delta} \right) \right) \\ &= \alpha \nu_{2\delta} \left((1 - \theta)^\delta - (1 + \theta)^\delta \right) \left(\alpha_0 \mathbb{E} \left(\rho_0^{2\delta} \right) + \beta \mathbb{E} \left(\rho_0^{3\delta} \right) \right) + \\ &\quad + \frac{\alpha^2}{2} \nu_{3\delta} \mathbb{E} \left(\rho_0^{3\delta} \right) \left((1 - \theta)^{2\delta} - (1 + \theta)^{2\delta} \right) \\ &= \frac{\alpha}{2} \mathbb{E} \left(\rho_0^{2\delta} \right) \nu_{2\delta} \left((1 - \theta)^\delta - (1 + \theta)^\delta \right) \times \\ &\quad \times \left(2\alpha_0 + \left(2\beta + \alpha \left((1 - \theta)^\delta + (1 + \theta)^\delta \right) \frac{\nu_{3\delta}}{\nu_{2\delta}} \right) \frac{\mathbb{E} \left(\rho_0^{3\delta} \right)}{\mathbb{E} \left(\rho_0^{2\delta} \right)} \right). \end{aligned}$$

We also observe that

$$\sqrt{\text{Var}(\rho^{2\delta})} = \sqrt{\text{Var}(\rho^\delta)} \frac{\text{cv}_2}{\text{cv}_1} \frac{\mathbb{E}(\rho^{2\delta})}{\mathbb{E}(\rho^\delta)}.$$

This combined with the first part completes the proof. \square

Remark 6. The coefficients of covariation that are presented in the above result and other moments of ρ_0 can be computed explicitly in terms of the model parameters using Proposition 6 and, in particular, Lemma 1, Lemma 2, Remark 2 and Corollary 2.

References

- [1] Jonas Andersson. On the normal inverse gaussian stochastic volatility model. *Journal of Business & Economic Statistics*, 19(1):44–54, 2001.

- [2] Mohamed El Babsiri and Jean-Michel Zakoian. Contemporaneous asymmetry in garch processes. *Journal of Econometrics*, 101:257–294, 2001.
- [3] F. Black. Studies in stock price volatility changes. In *Proceedings of the 1976 business meeting of the business and economics section*, pages 177–181. American Statistical Association, 1976.
- [4] T. Bollerslev. Generalized conditional autoregressive heteroskedasticity. *Journal of Econometrics*, 31:307–327, 1986.
- [5] 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.
- [6] R.F. Engle. Autoregressive conditional heteroskedasticity with estimates of the variance of u.k. inflation. *Econometrica*, 45:987–1007, 1982.
- [7] J. Eric, S. Poon, and M. Rockinger. *Financial Modelling Under Non-Gaussian Distributions*. Springer Finance, 2007.
- [8] B. Feunou and R. Tédongap. A stochastic volatility model with conditional skewness. *Journal of Business & Economic Statistics*, 30(4):576–591, 2012.
- [9] P. R. Hansen and A. Lunde. A forecast comparison of volatility models: does anything beat a garch(1,1)? *Journal of Applied Econometrics*, 20(7):873–889, 2005.
- [10] 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*, chapter 19, pages 462–474. Oxford University Press, 1999.
- [11] Changli He. *Statistical properties of GARCH processes*. PhD thesis, Stockholm School of Economics, 1997.
- [12] F. Javed and K. Podgórski. Leverage effect for volatility with generalized laplace error. *Economic Quality Control*, 29(2):157–166, 2014.
- [13] M. B. Jensen and A. Lunde. The NIG-S&ARCH model: A fat-tailed, stochastic, and autoregressive conditional heteroskedastic volatility model. *Econometrics Journal*, 4:167–342, 2001.
- [14] E. Jondeau and M. Rockinger. Conditional volatility, skewness, and kurtosis: existence, persistence, and comovements. *Journal of Economic Dynamics and Control*, 27(10):1699–1737, 2003.
- [15] B. B. Mandelbrot. The variation of certain speculative prices. *Journal of Business.*, 36:392–417, 1963.

- [16] M. McKenzie and H. Mitchell. Generalized asymmetric power ARCH modelling of exchange rate volatility. *Applied Financial Economics*, 12:555–564, 2002.
- [17] Gabriel Perez-Quiros and Allan Timmermann. Business cycle asymmetries in stock returns: Evidence from higher order moments and conditional densities. *Journal of Econometrics*, 103:259–306, 2001.
- [18] Nityananda Sarkar. Arch model with box–cox transformed dependent variable. *Stat. Prob. Lett.*, 50:365–374, 2000.
- [19] N. Shephard. Statistical aspects of arch and stochastic volatility. In O.E. Barndorff-Nielsen D.R. Cox, D.V. Hinkley, editor, *Time series models. In econometrics, finance and other fields*, pages 1–67. Chapman & Hall, London, 1995.
- [20] P. Verhoeven and M. McAleer. Fat tails and asymmetry in financial volatility models. *Mathematics and Computers in Simulation*, 64(3-4):351–361, 2004. MSSANZ/IMACS 14th Biennial Confernece on Modelling and Simulation.
- [21] Christina D. Wang and Per A. Mykland. The estimation of leverage effect with high-frequency data. *Journal of the American Statistical Association*, 109(505):197–215, 2014.