



ELSEVIER

Journal of Financial Economics 39 (1995) 71–104

JOURNAL OF
Financial
ECONOMICS

All in the family Nesting symmetric and asymmetric GARCH models

Ludger Hentschel

*William E. Simon Graduate School of Business Administration, University of Rochester,
Rochester, NY 14627, USA*

(Received April 1994; final version received November 1994)

Abstract

This paper develops a parametric family of models of generalized autoregressive heteroskedasticity (GARCH). The family nests the most popular symmetric and asymmetric GARCH models, thereby highlighting the relation between the models and their treatment of asymmetry. Furthermore, the structure permits nested tests of different types of asymmetry and functional forms. Daily U.S. stock return data reject all standard GARCH models in favor of a model in which, roughly speaking, the conditional standard deviation depends on the shifted absolute value of the shocks raised to the power three halves and past standard deviations.

Key words: GARCH; Asymmetry; Heteroskedasticity; Variance; Volatility

JEL classification: G12; C22

1. Introduction

This paper develops a family of models of generalized autoregressive heteroskedasticity (GARCH) that encompasses all the popular existing GARCH

I have benefited greatly from the advice and support of John Y. Campbell, conversations with Robin Lumsdaine, Pierre Perron, and Jim Powell, as well as comments by Jean-Michel Zakoian. In addition, the comments of Daniel Nelson (the referee) greatly improved the paper and in particular Appendix A. Last, but not least, I am grateful for the comments of many seminar participants. Of course, all remaining errors are solely my responsibility. An earlier version of this work was circulated under the title 'The Absolute Value GARCH Model and the Volatility of U.S. Stock Returns'. Financial support from the John M. Olin Foundation is gratefully acknowledged.

models. The nesting clearly shows the connection between the existing models, and permits new standard nested tests to determine the relative quality of each of the models' fits. The nested models include Bollerslev's (1986) GARCH model, Nelson's (1991) exponential GARCH (EGARCH) model, Zakoian's (1991) threshold GARCH (TGARCH) model, and others. The family of GARCH models is most easily derived from the asymmetric absolute value GARCH model based on the work by Taylor (1986), Schwert (1989), and Nelson (1992a). The absolute value GARCH model describes the conditional standard deviation as a linear combination of the absolute value of the shock and the lagged conditional standard deviation. Asymmetry is introduced by shifting and 'rotating' the absolute value of the shock. It is not difficult to derive a family of symmetric and asymmetric GARCH models by applying a Box-Cox (1964) transformation to the conditional standard deviation, and allowing different powers of the transformed shocks in the potentially asymmetric absolute value GARCH model.

Following the introduction of models of autoregressive conditional heteroskedasticity (ARCH) by Engle (1982) and their generalization by Bollerslev (1986), there have been numerous refinements of this approach to modeling conditional volatility. Most of the refinements have been driven by two empirical regularities of United States equity returns.

First, equity returns are strongly asymmetric: Negative returns are followed by larger increases in volatility than equally large positive returns. Following Black's (1976) exploration of this phenomenon, it is now commonly referred to as the 'leverage effect'. Black reasoned that large declines in equity values would not be matched by declines in the value of debt that is senior to equity, and would thereby raise the debt-to-equity ratio. This would increase the risk associated with the junior claim of equity. Black realized, however, that the financial leverage effect alone is empirically insufficient to explain the size of the observed asymmetry. This has also been documented by Christie (1982) and Schwert (1989). In the absence of a good theoretical model for this asymmetry, the GARCH literature has searched for econometric ways of describing the asymmetry. Models such as the EGARCH process introduced by Nelson (1991), the quadratic GARCH process of Sentana (1991) and Engle (1990), and the TGARCH model of Zakoian (1991) are among the popular asymmetric GARCH models.

A second empirical finding, one which precedes the GARCH literature (Mandelbrot, 1963), is that stock returns are fat-tailed. This leptokurtosis is reduced when returns are normalized by the time-varying variances of GARCH models, but it is by no means eliminated. Even allowing for changing variances, there remain too many very large events. These large events have caused concern about their effect on variance estimates from GARCH models. A standard GARCH model computes the next period's variance by squaring the current period's shock. For very large shocks, this produces dramatic increases in variance. Friedman and Laibson (1989) argue that large shocks constitute

extraordinary events and propose to truncate their influence on the conditional variance. Taylor (1986) and Schwert (1989) suggest a less drastic approach. They propose an ARCH model that specifies the conditional standard deviation as a moving average of lagged absolute residuals. Their argument follows Davidian and Carroll (1987), who show in a regression framework that variance estimators based on absolute residuals are robust to outliers. Nelson and Foster (1994) explicitly show that a similar argument can be made in a GARCH context. Neither the ARCH formulation of Taylor/Schwert nor the GARCH formulation of Nelson (1990a) permits asymmetries in the variance process, but both can be extended to account for the aforementioned empirical regularity.

These two stylized facts – asymmetry and leptokurtosis – seem to be responsible for the plethora of extant GARCH models. Unfortunately, the GARCH models themselves do not display obvious links to one another. The family of GARCH models proposed in this paper provides a unifying framework in which the models can be viewed and tested.

The character of the evolution of conditional variances is also of interest in the context of recent theoretical work on stochastic stock return variances. The common approach in this work has been to somewhat arbitrarily assume a particular diffusion process for the conditional variance. For efficient estimation under the null hypothesis that a particular diffusion model is correct, one should use a GARCH process that converges to the assumed diffusion limit.

Nelson (1990a) and Nelson and Foster (1994) derive the diffusion limits of the standard GARCH, the exponential GARCH, and the absolute value GARCH processes. The diffusion limit of the absolute value GARCH model is used in work by Scott (1987), and arises as a special case of Johnson and Shanno (1987) and Wiggins (1987). The diffusion limit of EGARCH is also a popular assumption, and is employed in Melino and Turnbull (1990) and the empirical work of Wiggins (1987). In contrast, the diffusion limit of the standard GARCH process is rarely used in theoretical work (although, see Hull and White, 1987).

Since little is known about the relative quality of fit of the various GARCH processes, the choice of the diffusion limit has thus far been guided more by convenience than realism. The tests based on the GARCH family should serve as a useful guide in choosing among the possible diffusion processes for the variance.

The remainder of this paper is organized as follows. Section 2 introduces the family of GARCH models and shows how its most important members are derived. Section 3 discusses the estimation of the GARCH models and gives results for U.S. stock returns. Section 4 concludes.

2. A family of variance models

This section establishes the asymmetric absolute value GARCH model. It then generalizes this model to a family of GARCH models that includes all of

the most popular symmetric and asymmetric GARCH models. *A priori*, the absolute value GARCH model holds no particular rank in this family; it is merely a convenient starting point from which the family of GARCH models can be established.

2.1. The absolute value GARCH model

Univariate GARCH models consist of two equations. The first, the mean equation, describes the observed data as a function of other variables plus an error term. The second, the variance equation, specifies the evolution of the conditional variance of the error from the mean equation as a function of past conditional variances and lagged errors. The specification of the mean equation is not without interest, but this paper is primarily concerned with the variance equation. In order to focus attention on the comparison of the different variance specifications, the same mean equation is used throughout.

2.1.1. The mean equation

One of the most common mean equations for excess returns on stocks is the GARCH-in-Mean (GARCH-M) model of Engle, Lilien, and Robins (1987). Mean equations of this form have been widely used in empirical studies of time-varying risk premia and the behavior of stock return variances (for example, Bollerslev, 1987; French, Schwert, and Stambaugh, 1987; Nelson, 1991; Sentana, 1991). They can be written as

$$r_{t+1} = \mu + \gamma \sigma_t^2 + \eta_{t+1}. \quad (1)$$

The logarithm of the return on the market portfolio in excess of the risk-free rate is denoted as r_{t+1} , and is given by a constant, μ , a time-varying risk premium, $\gamma \sigma_t^2$, and a heteroskedastic error term, η_{t+1} . The error has conditional variance σ_t^2 , which is known at time t . Conditional on the variance σ_t^2 , the error η_{t+1} can be decomposed as

$$\eta_{t+1} = \sigma_t \varepsilon_{t+1}, \quad (2)$$

where ε_{t+1} has zero mean and unit variance.

2.1.2. The variance equation

The common denominator of existing GARCH models is that they postulate that a transformation of the conditional standard deviation is linearly related to (nonnegative) functions of past and present shocks, plus a moving average of transformed standard deviations. In the case of Bollerslev's (1986) GARCH(p, q) model, the transformations for the shocks and the standard deviations are all

simple quadratics, so that

$$\sigma_t^2 = \omega + \sum_{i=1}^q \alpha_i \sigma_{t-i}^2 \varepsilon_{t+1-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2, \quad (3)$$

which reduces to Engle's (1982) ARCH(q) model when $p = 0$. All of the GARCH models that follow can be generalized to higher-order models or reduced to ARCH models. To simplify exposition, however, I will only present the GARCH(1,1) equivalents.

A GARCH(1,1) model for the standard deviation, σ_t , can be written as

$$\sigma_t = \omega + \alpha \sigma_{t-1} |\varepsilon_t| + \beta \sigma_{t-1}. \quad (4)$$

As is readily apparent, this formulation is identical to a GARCH(1,1) process, except that the conditional variance terms, σ_t^2 , have been replaced by their square roots, and the squared error has been replaced by its absolute value.

Taylor (1986) and Schwert (1989) first suggested ARCH models for the conditional standard deviation. In these models, the conditional variance is the square of a weighted average of absolute shocks, not the weighted average of squared shocks. Because of Jensen's inequality, large shocks have a smaller effect on the conditional variance than in the standard GARCH model. This intuition is supported by Nelson and Foster (1994). Nelson and Foster show that a GARCH extension of the Taylor/Schwert ARCH model is a consistent estimator of the conditional variance of near diffusion processes; furthermore, they show that in the presence of leptokurtic error distributions, the absolute value GARCH model is a more efficient filter of the conditional variance than Bollerslev's (1986) GARCH.

2.1.3. Modeling asymmetry

Although Eq. (4) is a well-specified variance equation, the standard deviation's symmetric response to shocks ignores the empirically important asymmetries discussed in the introduction. This section therefore generalizes the model to permit an asymmetric response.

The asymmetric version of the absolute value GARCH model is given by

$$\sigma_t = \omega + \alpha \sigma_{t-1} f(\varepsilon_t) + \beta \sigma_{t-1}, \quad (5)$$

where

$$f(\varepsilon_t) = |\varepsilon_t - b| - c(\varepsilon_t - b). \quad (6)$$

In the interest of brevity, I refer to this absolute value GARCH equation as the AGARCH(1,1) model.¹

¹Engle and Ng (1993) named a variance equation that specifies the conditional variance as a function of a shifted parabola the AGARCH model. This model has been more fully worked out by Sentana (1991) who refers to it as QGARCH. I hope that this duplication of acronyms does not lead to confusion.

The ‘news impact curve’ introduced by Pagan and Schwert (1990), and so christened by Engle and Ng (1993), is a useful tool in discussing asymmetry. The news impact curve relates revisions in conditional volatility to shocks. In the context of the absolute value GARCH model, it is convenient to investigate the impact of shocks on the conditional standard deviation. As Fig. 1a shows, the news impact curve of Eq. (4) is symmetric in $\varepsilon_t - \sigma_t$ space.

The AGARCH model incorporates both of the previously used main approaches to permit asymmetry in the variance equation. The first approach employs shifts in the news impact curve. In the AGARCH model, the magnitude and direction of such a shift are controlled by the parameter b ; a positive value of b causes a rightward shift of the news impact curve. If, as in Fig. 1b, the news impact curve is shifted to the right by the distance b , one obtains asymmetry that matches the stylized facts of stock return volatility: For negative shocks, volatility rises more than for equally large, but positive, shocks. The quadratic GARCH model of Sentana (1991) and Engle (1990), and the nonlinear–asymmetric

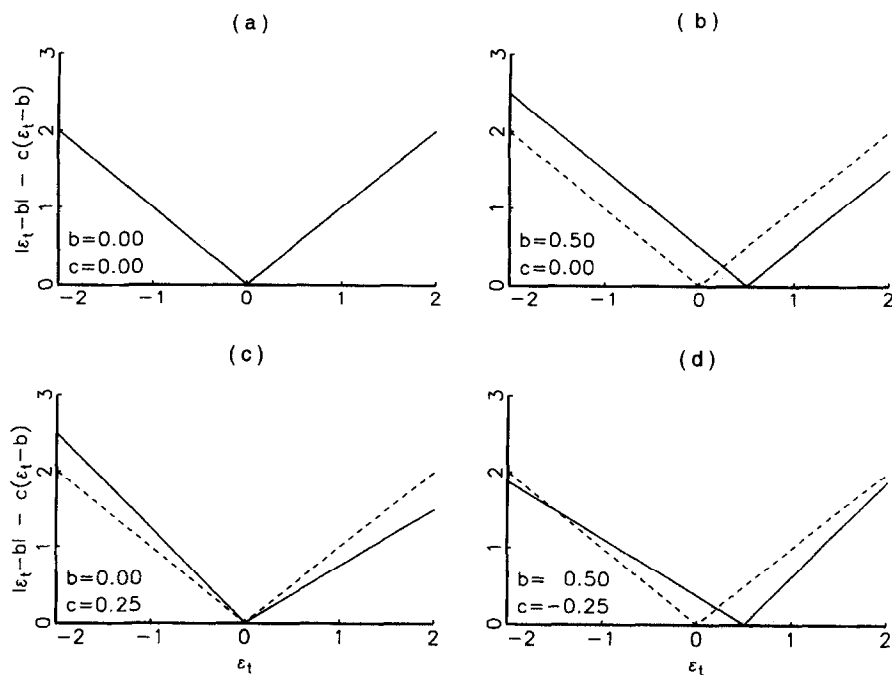


Fig. 1. The asymmetric transformation $f(\varepsilon_t)$.

The panels show the shifted and rotated absolute value function $f(\varepsilon_t) = |\varepsilon_t - b| - c(\varepsilon_t - b)$ [Eq. (6) in text], for various shifts and rotations, b and c . The dashed line shows the absolute value function, $|\varepsilon_t|$, for comparison. The transformation $f(\cdot)$ controls the effect of shocks, ε_t , on the conditional volatility, σ_t , in the AGARCH model $\sigma_t = \omega + \alpha\sigma_{t-1}f(\varepsilon_t) + \beta\sigma_{t-1}$ [Eq. (5) in text].

ARCH model of Engle and Ng (1993) employ a shifted news impact curve to achieve asymmetry.

The second approach employs what may best be described as a rotation of the news impact curve. By allowing slopes of different magnitudes on either side of the origin, the news impact curves of this type also produce asymmetric variance responses. In the AGARCH model, the rotation is governed by the parameter c ; a positive value of c corresponds to a clockwise rotation. If the news impact curve is rotated clockwise, like the curve shown in Fig. 1c, negative shocks increase volatility more than positive shocks. The EGARCH model of Nelson (1991) and the models by Glosten, Jagannathan, and Runkle (1993) and Zakoian (1991) feature rotated news impact curves. Note that c does not cause a pure rotation of the absolute value function. Rather, c controls the slopes of the news impact curve, which are different on either side of the minimum at $\varepsilon_t = b$. To achieve a pure rotation, one must also pick an appropriate value of α .

Current GARCH models only permit either a shift or a rotation, but not both. In principle, these two types of asymmetry are distinct, and should not be treated as substitutes for each other. As can be seen from Fig. 1b, the asymmetry caused by the shift is most pronounced for small shocks. For extremely large shocks, the asymmetric effect becomes a negligible part of the total response. On the other hand, the rotated news impact curve of Fig. 1c maintains the hypothesis that a zero shock results in the smallest increase of conditional variance. Additionally, the size of the asymmetric effect of small shocks is very small in absolute terms. The size of the asymmetric effect relative to the total response, however, is constant.

When the shift and rotation are combined in one news impact curve, they can either reinforce or offset each other. The latter, less-obvious case is depicted in Fig. 1d. By appropriately shifting and rotating the news impact curve, it is possible to have asymmetry for small shocks, a roughly symmetric response for moderate shocks, and asymmetry for very large shocks. The asymmetry for the large shocks, however, is the opposite of the asymmetry for the small shocks. This is true because the shift is the dominant source of asymmetry for small shocks, while the rotation is more important for large shocks.

2.2. Other family members

The AGARCH model of the previous section is interesting in its own right, but it is only one member of a large family of variance models. This section derives this family of GARCH models and points out its best-known members.

By rewriting the intercept in Eq. (5) and introducing the parameters λ and ν , one can, for $\lambda = \nu = 1$, write Eq. (5) as

$$\frac{\sigma_t^\lambda - 1}{\lambda} = \omega' + \alpha \sigma_{t-1}^\lambda f^\nu(\varepsilon_t) + \beta \frac{\sigma_{t-1}^\lambda - 1}{\lambda}. \quad (7)$$

In general, this is a law of motion for the Box–Cox (1964) transformation of the conditional standard deviation, and the parameter λ determines the shape of the transformation. For $\lambda > 1$ the transformation of σ_t is convex, while for $\lambda < 1$ it is concave. The parameter ν serves to transform the (potentially shifted and rotated) absolute value function $f(\cdot)$. Fig. 2 shows that although the transformation is asymmetric, convex functions of ε_t drive the variance equation for $\nu \geq 1$. For $0 < \nu < 1$, on the other hand, the transformation is concave on either side of b . By construction, the AGARCH model corresponds to the case of $\lambda = \nu = 1$.

The remainder of this section demonstrates that Eq. (7) nests all popular GARCH models. The special cases are obtained by appropriately choosing the parameters λ , ν , b , and c . The nesting results of this section and the appropriate restrictions on λ , ν , b , and c are summarized in Table 1. The restrictions that either b or c or both are zero are unnecessary in all of the models, but are given in Table 1 as stated by the original authors.

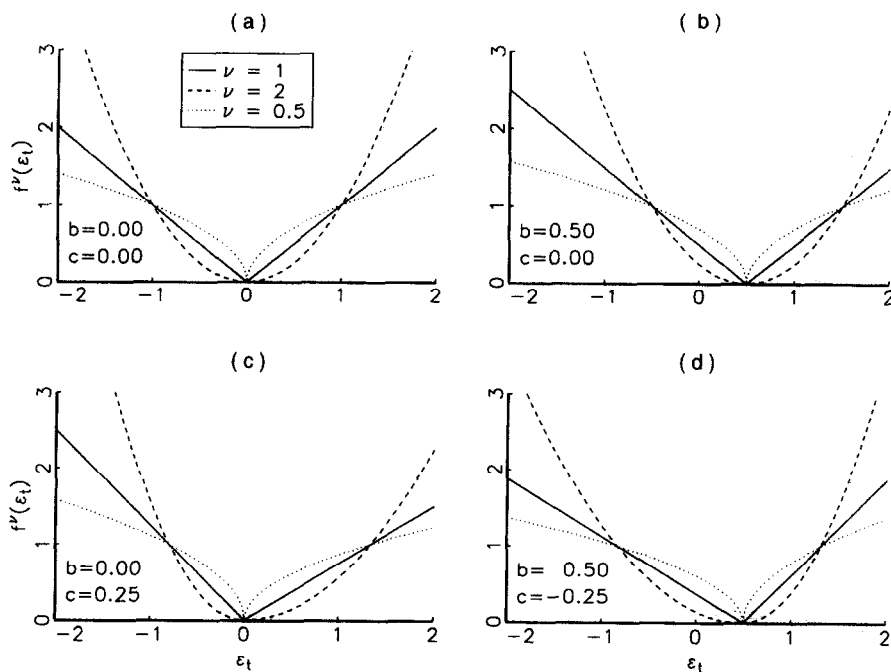


Fig. 2. The transformation $f^\nu(\varepsilon_t)$.

The panels show the transformation $f^\nu(\varepsilon_t)$ for $f(\varepsilon_t) = |\varepsilon_t - b| - c(\varepsilon_t - b)$ [Eq. (6) in text], and different values of ν , b , and c . The transformation $f^\nu(\cdot)$ controls the impact of shocks, ε_t , on the transformed conditional volatility, σ_t , in the variance equation $(\sigma_t^2 - 1)/\lambda = \omega' + \alpha \sigma_{t-1}^\lambda f^\nu(\varepsilon_t) + \beta ((\sigma_{t-1}^\lambda - 1)/\lambda)$ [Eq. (7) in text].

Table 1
Nested GARCH models

λ	v	b	c	Model
0	1	0	free	Exponential GARCH (Nelson)
1	1	0	$ c \leq 1$	Threshold GARCH (Zakoian)
1	1	free	$ c \leq 1$	Absolute value GARCH (Taylor/Schwert)
2	2	0	0	GARCH (Bollerslev)
2	2	free	0	Nonlinear-asymmetric GARCH (Engle, Ng)
2	2	0	free	GJR GARCH (Glosten, Jagannathan, Runkle)
free	λ	0	0	Nonlinear ARCH (Higgins, Bera) ^a
free	λ	0	$ c \leq 1$	Asymmetric power ARCH (Ding, Granger, Engle) ^a

The first four columns list the restrictions one has to apply to the variance equation $(\sigma_t^2 - 1)/\lambda = \omega + \alpha\sigma_{t-1}^2 f^v(\varepsilon_t) + \beta((\sigma_{t-1}^2 - 1)/\lambda)$ [Eq. (7) in text], $f(\varepsilon_t) = |\varepsilon_t - b| - c(\varepsilon_t - b)$ [Eq. (6) in text], in order to obtain the model named in the last column. In all members of this family of GARCH models, a transformation of the conditional standard deviation, σ_t , is determined by the transformation $f(\cdot)$ of the innovations, ε_t , and lagged transformed conditional standard deviations.

^a Nested if $\lambda \neq 0$.

All conditional variance models should ensure that σ_t^2 takes real, positive values. In the case of Bollerslev's (1986) GARCH model, this requirement implies that ω , α , and β have to be positive. These conditions are sufficient, but not always necessary, for positive conditional variances in other models in the family.² Additionally, to ensure the positivity of σ_t^2 , many models require that $f^v(\varepsilon_t)$ is positive. Positivity of $f^v(\varepsilon_t)$ is guaranteed when $|c| \leq 1$, which ensures that neither arm of the rotated absolute value function crosses the abscissa. The parameter b , however, is unrestricted in size and sign. For all the models in the family, there are restrictions on the magnitude of the parameters to ensure covariance stationarity of the process. All of the restrictions for positivity and stationarity are discussed in Appendix A. Typically, none of these restrictions are binding. Consequently, they are not an important factor in model choice, and have been suppressed in Table 1.

2.2.1. Exponential GARCH ($\lambda = 0$, $v = 1$, $b = 0$)

Using l'Hôpital's rule, one can show that the Box-Cox transformation converges to the natural logarithm as λ goes to zero: $\lim_{\lambda \rightarrow 0} (\sigma_t^2 - 1)/\lambda = \ln \sigma_t$.

From l'Hôpital's rule it is clear that for $v = 1$, Eq. (7) converges to the exponential GARCH model introduced by Nelson (1991) as λ goes to zero. If b is

²Slightly negative values of α_i and β_j for higher-order lags don't result in negative conditional volatility in any of these models. See Nelson and Cao (1992) for necessary restrictions on α_i and β_j in the standard GARCH(p, q) model with $p = \{0, 1, 2\}$.

set to zero, and the constant, unconditional mean of $f(\varepsilon_t)$ is subtracted from $f(\varepsilon_t)$ and added to the intercept, the variance equation becomes

$$\ln \sigma_t^2 = 2\omega'' + 2\alpha[|\varepsilon_t| - E|\varepsilon_t| - c\varepsilon_t] + \beta \ln \sigma_{t-1}^2, \quad (8)$$

which is identical to Nelson's EGARCH. Since exponentiation also ensures positivity, EGARCH does not impose sign restrictions on ω'' , α , and β . More generally, Eq. (7) allows not only for the rotation of the news impact curve given in the standard formulation of Nelson's EGARCH, but also for the shift that was discussed in the previous section. This can be seen by substituting $(\varepsilon_t - b)$ for ε_t in Eq. (8).

Note that it is crucial that $f(\varepsilon_t)$ is not raised to the power λ in order to nest EGARCH. By raising $f(\varepsilon_t)$ to the power λ , one obtains a version of the asymmetric power ARCH model of Ding, Granger, and Engle (1993) or the nonlinear ARCH model of Higgins and Bera (1992), which are discussed later.

2.2.2. Threshold GARCH ($\lambda = \nu = 1$, $b = 0$)

Like the AGARCH model, Zakoian's (1991) threshold GARCH model also treats the conditional standard deviation as a linear function of shocks and lagged standard deviations. By setting $\lambda = \nu = 1$ and permitting c to enter the variance equation, one can transform Eq. (7) into the standard TGARCH notation

$$\begin{aligned} \sigma_t &= \omega + \alpha \sigma_{t-1} [|\varepsilon_t| - c\varepsilon_t] + \beta \sigma_{t-1} \\ &= \omega + \alpha(1 - c)\sigma_{t-1}\varepsilon_t^+ - \alpha(1 + c)\sigma_{t-1}\varepsilon_t^- + \beta \sigma_{t-1}, \end{aligned} \quad (9)$$

where ε_t^+ stands for $\max\{\varepsilon_t, 0\}$ and ε_t^- for $\min\{\varepsilon_t, 0\}$.

Although TGARCH permits a rotation of the news impact curve, it does not allow a shift. More recently, Rabemananjara and Zakoian (1993) have proposed a class of threshold GARCH models that permits σ_t to become negative, arguing that the conditional variance σ_t^2 is still positive. When the news impact curves from these models are plotted in $\varepsilon_t - |\sigma_t|$ space, the negative values of σ_t are reflected at the x-axis, and a type of asymmetry not unlike the shift of the AGARCH news impact curve can be achieved. In this TGARCH model, however, the 'shift' and tilt are inextricably linked. One could introduce an independent shift by substituting $(\varepsilon_t - b)$ for ε_t in the equations above, which can be captured by reinterpreting ε_t^+ as $\max\{\varepsilon_t - b, 0\}$ and ε_t^- as $\min\{\varepsilon_t - b, 0\}$.

2.2.3. Standard GARCH ($\lambda = \nu = 2$, $b = c = 0$)

The GARCH model of Bollerslev (1986) can be obtained by restricting $f(\varepsilon_t)$ to be the simple absolute value, $|\varepsilon_t|$, and setting both λ and ν to 2:

$$\sigma_t^2 = 2\omega'' + 2\alpha\sigma_{t-1}^2\varepsilon_t^2 + \beta\sigma_{t-1}^2. \quad (10)$$

This model has become the most commonly used, and therefore ‘standard’, GARCH model. Nonetheless, because of the restrictions on b and c , it does not allow for any asymmetries in the response of the conditional variance. The next two models are standard GARCH models, except that one allows for a shift and the other for a rotation of the news impact curve.

2.2.4. Nonlinear–asymmetric GARCH ($\lambda = \nu = 2, c = 0$)

When $\lambda = \nu = 2$ and $c = 0$, but b is freely estimated, one obtains the nonlinear–asymmetric GARCH of Engle and Ng (1993), where

$$\sigma_t^2 = 2\omega'' + 2\alpha\sigma_{t-1}^2(\varepsilon_t - b)^2 + \beta\sigma_{t-1}^2. \quad (11)$$

In this model, the news impact curve is shifted to the right by the distance b .

2.2.5. Glosten–Jagannathan–Runkle GARCH ($\lambda = \nu = 2, b = 0$)

When $f(\varepsilon_t)$ is the rotated absolute value function with $b = 0$ but a free c , then the model reduces to the Glosten–Jagannathan–Runkle (1993) (GJR) GARCH model for $\lambda = \nu = 2$:

$$\sigma_t^2 = \omega'' + 2\alpha\sigma_{t-1}^2[(1 + c^2)\varepsilon_t^2 - 2c|\varepsilon_t|\varepsilon_t] + \beta\sigma_{t-1}^2. \quad (12)$$

The news impact curve for this model, although quadratic, has different slopes on either side of the origin. This can be seen more directly by rewriting Eq. (12) as

$$\sigma_t^2 = \begin{cases} \omega'' + 2\alpha(1 + c)^2\sigma_{t-1}^2\varepsilon_t^2 + \beta\sigma_{t-1}^2, & \varepsilon_t < 0, \\ \omega'' + 2\alpha(1 - c)^2\sigma_{t-1}^2\varepsilon_t^2 + \beta\sigma_{t-1}^2, & \varepsilon_t \geq 0. \end{cases} \quad (13)$$

If $c > 0$, the news impact curve is steeper for negative shocks than it is for positive shocks.

Thus, the standard GARCH model, the nonlinear–asymmetric GARCH model, and the GJR GARCH model only differ in their restrictions on b and c . Indeed, a model that permits both a shift and a rotation of the news impact curve can be obtained by permitting nonzero values for both b and c .

2.2.6. Nonlinear ARCH ($\lambda = \nu, b = c = 0$)

Higgins and Bera (1992) also draw on the Box–Cox transformation to nest symmetric ARCH models. They call the process ‘nonlinear ARCH (NARCH)’. Higgins and Bera only provide an ARCH specification, but a GARCH extension of their process is

$$\sigma_t^\lambda = \omega'' + \alpha\lambda|\sigma_{t-1}\varepsilon_t|^\lambda + \beta\sigma_{t-1}^\lambda. \quad (14)$$

Even for $\beta = 0$, the process in Eq. (7) is not quite identical to the NARCH model. For $\lambda = 0$, NARCH converges to a specification suggested by Geweke (1986) and Pantula (1986), in which the natural logarithm of the standard deviation

depends on the logarithm of the absolute value of the errors and the logarithm of the lagged standard deviation. Unfortunately, such a specification breaks down for any $\varepsilon_t = 0$, and poses numerical difficulties for ε_t very close to zero. [Nevertheless, a model in which the log of the normalized shock drives the variance could be nested by employing the Box–Cox transformation of $f(\varepsilon_t)$ instead of $f^v(\varepsilon_t)$.] The model introduced here becomes deterministic for $\lambda = v = 0$. In that case, the conditional standard deviation converges to and remains at its unconditional mean. Empirically, only values of $\lambda > 0$ appear to be of relevance and the model in Eq. (7) nests NARCH for all nonzero values of λ and v .

2.2.7. Asymmetric power ARCH ($\lambda = v$, $b = 0$)

Based on the nonlinear GARCH model, Ding, Granger, and Engle (1993) developed a family of asymmetric GARCH models. They call this model asymmetric power ARCH (A-PARCH). A-PARCH generalizes the model in Eq. (14) to include a rotation of the news impact curve

$$\sigma_t^\lambda = \omega'' + \alpha\lambda(|\sigma_{t-1}\varepsilon_t| - c\sigma_{t-1}\varepsilon_t)^\lambda + \beta\sigma_{t-1}^\lambda, \quad (15)$$

but the process is otherwise identical to the nonlinear GARCH model.

2.2.8. Other related models

Another class of GARCH models is obtained by setting $\lambda = 2$ and $v = 1$. These models are similar to processes proposed by Engle (1982), for which the conditional variance evolves as a function of lagged absolute residuals. In the Eq. (7) model, however, the absolute shocks would be scaled by their variance, not their standard deviation. For these processes, the conditional variance responds even less to large shocks than in the absolute value GARCH model, and by choosing larger values of λ , this property can be strengthened further. Alternatively, one can reduce v to flatten the response to large shocks.

Symmetric models with $v < 1$ are not unlike Friedman and Laibson's (1989) model. They argue that large shocks should exert relatively less influence on the conditional variance. Friedman and Laibson's (1989) variance equation truncates the effect of large shocks by taking the sine of the scaled, squared residuals and smoothly pasting it to the constant 1 at $\pi/2$. Fig. 2 shows that $f^v(\varepsilon_t)$ is locally concave for $0 < v < 1$, just like the sine function, but unlike the pasted sine function, $f^v(\varepsilon_t)$ does not level off completely.

2.3. Restrictions on v

The introduction of the second parameter, v , may seem artificial in view of the fact that most models set $\lambda = v$. Nonetheless, it is crucial that $\lambda \neq v$ in order to nest EGARCH. It is not difficult to specify an exponent for $f(\varepsilon_t)$ that is a function of λ and has the desired values of $\{1, 1, 2\}$ for corresponding values of λ of $\{0, 1, 2\}$. Such a model nests all of the common GARCH models, but avoids

the introduction of the second parameter. [The quadratic, $v = \lambda(\lambda - 1)/2 + 1$, is an obvious choice. Another, perhaps better, choice is $v = \max\{1, \lambda\}$.] Since such a formulation also restricts combinations of λ and v at other points, and since there is no compelling reason to believe that such a restriction reflects the truth, I retain the more general parameterization of Eq. (7), which employs v as a separate parameter.

2.4. Excluded models

The family of variance functions described by Eq. (7) is remarkably rich, but still excludes certain GARCH models.

In particular, the quadratic GARCH model of Sentana (1991) cannot be nested in this framework. Here, all shifts occur for the function of the normalized residuals. In the quadratic GARCH model, the residual itself is shifted by a fixed amount, regardless of the conditional variance. Evidence from Engle and Ng (1993) suggests that there is little empirical difference between the nonlinear–asymmetric GARCH, which shifts the normalized residuals by b and can be nested here, and the corresponding quadratic GARCH model. The most general cases of the quadratic GARCH model permit multiplicative interactions among lagged shocks. They cannot be directly compared to the nonlinear–asymmetric GARCH model, nor to any of the other GARCH models presented here.

2.5. Hypothesis testing

The nesting permits simple nested tests of model fit. If one can estimate the unrestricted GARCH model in Eq. (7), then tests of the quality of fit of any of the nested GARCH models are reduced to testing whether linear combinations of the parameters b , c , λ , and v are significantly different from zero. Asymptotic versions of such tests are easily constructed from the covariance matrix of the parameter estimates. Alternatively, one can conduct likelihood ratio tests of the parameter restrictions.

A potential problem arises from tests of models that impose sign restrictions on ω , α , and β (for example, GARCH) against models that don't impose such restrictions (for example, EGARCH). If EGARCH is estimated with negative parameters, tests on λ and v are no longer the appropriate test against the more restricted models in the family, since these have to have positive ω , α , and β . Yet, if one estimates a model under the restriction that ω , α , and β are positive, and if one can reject the alternative hypothesis of EGARCH, then this rejection will hold *a fortiori* due to the extra degrees of freedom in EGARCH. The same argument can be applied to the size restriction on c . In practice, the parameter estimates generically satisfy the sign and size restrictions, and these concerns are not important. These difficulties don't arise for λ and v , since these parameters don't have to satisfy similar restrictions.

One can, in principle, perform nonnested tests of one model against another. Unfortunately, these tests may well reject both models. Nonnested tests can in theory – and frequently do in practice – reject all models under consideration. The tests proposed here can reject any of the special cases I have discussed, but the rejections will always be in favor of a less-restricted alternative model.

3. Data, estimation, and results

Having derived the absolute value GARCH model and the associated family of conditional variance processes, I now apply the family of GARCH-M models to daily excess returns on U.S. equities.

3.1. Data

The daily stock returns are a combination of the Schwert (1990) data and the daily index files from the Center for Research in Security Prices (CRSP). The stock returns are for a value-weighted equity index and include dividends.³ The risk-free interest rate data were taken from the Fama/Bliss risk-free rate file on the CRSP bond tapes. They were converted to daily returns by assuming that the total annual return accrues continuously in calendar time. The data span the period from January 2, 1926 to December 31, 1990, and contain 17,486 observations. The difference between the log stock return and the log of the risk-free rate forms the log excess return, r_{t+1} , used in the analysis.

3.2. Estimation

Since the conditional variance appears directly in the mean equation, Eq. (1), the information matrix for the full model is not block-diagonal for the parameters of the mean and variance equations. Under these circumstances, efficient estimation requires numerical maximum likelihood estimation of the full model. In order to carry out the estimation, one needs to make a distributional assumption for the error term, ε_{t+1} . In light of the empirical evidence of fat-tailed errors, several authors (Bollerslev, 1987; Nelson, 1991; Sentana, 1991) have chosen leptokurtic distributions such as the Student- t distribution or the

³From January 2, 1926 to January 3, 1928, the returns are computed from the Dow Jones indices for industrial and railroad stocks which are not value-weighted. From January 4, 1928 to July 2, 1962, the returns are computed from the S&P 500 industrials index. For both periods, dividend yields were computed by spreading the dividend yields from the monthly CRSP files evenly across the trading days in each month. From July 3, 1962 to December 31, 1990, the data are the value-weighted CRSP returns including dividends.

Generalized Error Distribution (GED). For simplicity, I assume that the error term is drawn from a normal density

$$\varepsilon_{t+1} \sim N(0, 1). \quad (16)$$

In the absence of normally distributed errors, this procedure can be viewed as quasi-maximum likelihood in the sense of White (1982). Fat-tailed distributions typically achieve a somewhat better fit than the normal density. Even so, they generally do not change any of the results – with the obvious exception of the likelihood value. In particular, the parameter estimates and associated standard errors are little changed, and likelihood ratio tests of parameter restrictions yield answers that are very similar to those obtained under the assumption of normality.⁴ Even these fat-tailed models, however, typically reject the hypothesis that the estimated errors were drawn from the assumed leptokurtic distributions.

All models were estimated using numerical maximum likelihood. There are currently no known sufficient conditions for consistency of this procedure in the case of GARCH-M models. Instead, the literature has proceeded without this knowledge, under the reasonable belief that such conditions will eventually be found. A similar caveat applies to the asymptotic normality of the maximum likelihood estimators of GARCH-M models. Recent work by Lumsdaine (1992) has demonstrated asymptotic consistency and normality for a class of GARCH and integrated GARCH models, but these results have not yet been generalized to the GARCH-M case. Nonetheless, I compute standard errors according to the quasi-maximum likelihood procedure of Bollerslev and Wooldridge (1992). I follow the GARCH literature in interpreting these standard errors and other test statistics as though the estimators were indeed asymptotically normal.

The nondifferentiability of the absolute value function can pose minor numerical difficulties. These difficulties are easily overcome by approximating the absolute value of the shocks by a hyperbola, $|\varepsilon_t| \approx \sqrt{a^2 + \varepsilon_t^2}$. As a goes to zero, the approximation converges to the absolute value. Even for moderate values of a , the approximation is excellent and maintains differentiability. Appendix B shows that the first and second moments of the hyperbolic approximation of the potentially shifted and rotated absolute value function are within a few parts per million of the exact values when $a = 0.001$.

3.3. Results

Table 2 exhibits estimates of the variance equation parameters for several members of the GARCH model family. The estimates of the mean equation

⁴This statement was confirmed by estimating the AGARCH model under the assumption of Student- t distributed errors. The estimates indicated a Student- t distribution with about seven degrees of freedom, but were otherwise similar to those presented below.

Table 2
Estimates of variance equation parameters

Model	λ (SE)	ν (SE)	ω (SE)	α (SE)	β (SE)	b (SE)	c (SE)
$\ln \sigma_t^2 = \omega + \alpha[f(\varepsilon_t) - E[f(\varepsilon_t)]] + \beta \ln \sigma_{t-1}^2$							
EGARCH	0	1	-1.221×10^{-1} (0.186 $\times 10^{-1}$)	0.168 (0.015)	0.986 (0.020)	0.000	0.360 (0.045)
	0	1	-1.473×10^{-1} (0.213 $\times 10^{-1}$)	0.192 (0.017)	0.984 (0.023)	0.322 (0.021)	0.110 (0.047)
$\sigma_t = \omega + \alpha \sigma_{t-1} f(\varepsilon_t) + \beta \sigma_{t-1}$							
TGARCH	1	1	1.314×10^{-4} (0.239 $\times 10^{-4}$)	0.091 (0.010)	0.917 (0.009)	0.000	0.390 (0.046)
AGARCH	1	1	1.354×10^{-4} (0.247 $\times 10^{-4}$)	0.092 (0.010)	0.915 (0.009)	0.033 (0.017)	0.367 (0.051)
$\sigma_t^2 = \omega + \alpha \sigma_{t-1}^2 f^2(\varepsilon_t) + \beta \sigma_{t-1}^2$							
GARCH	2	2	8.573×10^{-7} (1.706 $\times 10^{-7}$)	0.094 (0.011)	0.902 (0.010)	0.000	0.000
N-A GARCH	2	2	10.909×10^{-7} (2.258 $\times 10^{-7}$)	0.087 (0.010)	0.887 (0.013)	0.474 (0.056)	0.000
GJR GARCH	2	2	9.412×10^{-7} (1.955 $\times 10^{-7}$)	0.078 (0.009)	0.909 (0.010)	0.000	0.255 (0.031)
	2	2	10.906×10^{-7} (2.256 $\times 10^{-7}$)	0.088 (0.011)	0.886 (0.014)	0.488 (0.087)	0.011 (0.047)

parameters are suppressed in the table, since they are not the focus of this study. The estimates are similar across all models, however. For the models shown, estimates of μ vary from 2.927×10^{-4} to 4.148×10^{-4} , with standard errors of about 0.8×10^{-4} . Estimates of γ range from -1.379 to 1.670 , but are almost always within two standard errors of zero.

The table shows five panels. The top three panels contain parameter estimates for EGARCH, AGARCH, and GARCH models. The fourth panel contains estimates for a model in which λ and ν have been restricted to equal one another. This class of models nests the GARCH and AGARCH families, but not the EGARCH models. The bottom panel shows the parameter estimates for the most general member of the family, a model in which all parameters are freely estimated.⁵

The order of the models in the first four panels is the same as in Table 1. The first column indicates the name of the model where applicable, and the second and third columns show values of λ and ν . In the first three panels, these parameters were restricted to the stated values according to each class of model. The last two panels show estimates of λ , and λ and ν , respectively. The asymptotic standard errors of all parameter estimates are shown in parentheses underneath the estimates. The fourth column shows that the scale parameter ω is significantly different from zero for all models. Of course, the magnitude of ω differs from model to model, concomitant with the magnitude of the volatility measure, $\ln \sigma_t^2$, σ_t , σ_t^2 , or σ_t^λ . The parameters α and β in the fifth and sixth columns are highly significant and of similar magnitude across models. Although the sum of α and β exceeds one for some of the models, none of the estimates violate the appropriate stationarity conditions given in Appendix A. In all five panels, the asymmetry of the variance equation is controlled by the parameters b and c , whose estimates are shown in the seventh and eighth columns, respectively.

The top panel shows that the standard EGARCH model is estimated with a highly significant rotation parameter, $\hat{c} = 0.360$. The magnitude of the rotation is reduced by the introduction of the shift in the second row of the panel, but it remains significant. Indeed, both the rotation and the shift are significantly different from zero.

The second panel shows the parameter estimates for the TGARCH and AGARCH models. The top row shows that the estimated rotation, \hat{c} , of the news impact curve is similar to that in the EGARCH model. Furthermore, the estimate is highly significant. But, unlike in the EGARCH model, the rotation

⁵The table only shows a subset of the models that were estimated. For each panel, models with the restrictions $b = c = 0$, $b = 0$, and $c = 0$ were estimated in addition to the models that freely estimate b and c . In the interest of brevity, not all estimates are shown or discussed. A complete listing of all parameter estimates is available upon request.

dominates the shift. When b and c are jointly estimated, b is small and insignificantly different from zero, while c is largely unchanged by the introduction of the shift parameter and remains highly significant.

The estimates for the GARCH models in the third panel reveal that all three of the asymmetric models easily reject the standard, symmetric GARCH model. In each of the asymmetric models, either b or c is significantly different from zero. When both a shift and a rotation of the news impact curve are estimated, the shift is to the right, while the rotation is counterclockwise but insignificantly different from zero. The rotation is the dominant effect for very large shocks. However, the point estimate of c implies that only negative shocks larger than 44 standard deviations result in smaller conditional volatility than equally large positive shocks. Hence, c is neither statistically nor economically significant in this model.

The fourth panel shows parameter estimates for models in which λ is freely estimated, but ν is restricted to equal λ . For all models in this class, the estimates of λ are close to 1.5. The standard errors show that λ is significantly different from zero, but, more interestingly, also different from one or two. Rejections of the latter two hypotheses constitute rejections of the AGARCH and standard GARCH models. As the estimates for b and c reveal, the shift is significantly different from zero, but the rotation is not.

The bottom panel shows the parameter estimates for the most general EGARCH-M(1, 1) model of the family. Here, the parameters λ , ν , b , and c are freely estimated. The exponent λ is estimated to be 1.131, which is significantly different from zero and two, but not from one. The exponent ν is estimated to be 1.524, and is significantly different from zero, one, or two. The standard errors for the two parameters are insufficient, however, to test the hypotheses that the two exponents are equal to one another, or that $\lambda = 0$ and $\nu = 1$. These hypotheses are tested later with likelihood ratio tests. Once again, the shift parameter b is significant, while the rotation parameter c is not.

In order to interpret the estimates of λ and ν , it is easiest to approximate them as $\lambda \approx 1$ and $\nu \approx 1.5$. Squaring the variance equation reveals that the conditional variance is a linear combination of the transformed shock, $f(\varepsilon_t)$, raised to the power 1.5 as well as raised to the power 3. The cubic term has a weight of $\alpha^2 \sigma_{t-1}^2$, and therefore is typically very small. The shock raised to the power 1.5, on the other hand, has a weight of $2\alpha \sigma_{t-1}(\omega + \beta \sigma_{t-1})$, which is much larger. Nevertheless, as ε_t becomes very large, the relation becomes approximately cubic.

The estimates of b and c imply that negative shocks result in higher volatility than equally large positive shocks, which is in accordance with the 'leverage effect'. Table 3 verifies that asymmetry is an important feature of daily U.S. stock returns. The first column lists the restrictions on λ and ν that form the maintained hypothesis for each panel. The second column shows the three of four possible types of asymmetry that form the null hypotheses for each model. If

there were no asymmetry, then b and c would equal zero, as indicated in the first row. This possibility can be tested against the three alternative hypotheses listed in the last three columns of the table. The asymmetry could be caused by a shift of the news impact curve, shown in the third column; it could be caused by a rotation of the news impact curve, shown in the fourth column; or it could be caused by both a shift and a rotation, shown in the fifth column. The likelihood ratio statistics have a χ^2 distribution with one or two degrees of freedom. The test statistics in the top row of each panel clearly exceed 200, and easily reject the possibility of a symmetric news impact curve in favor of any type of asymmetry. The probability levels, shown in parentheses underneath the test statistics, are

Table 3
Likelihood ratio tests for asymmetry in volatility

Maintained hypothesis	H_0	H_A		
		$c = 0, b$ free	$b = 0, c$ free	b and c free
$\lambda = 0, \nu = 1$	$b = c = 0$	303.622 (< 0.001)	304.542 (< 0.001)	311.667 (< 0.001)
	$c = 0, b$ free			8.045 (0.005)
	$b = 0, c$ free			7.125 (0.008)
$\lambda = 1, \nu = 1$	$b = c = 0$	315.098 (< 0.001)	335.082 (< 0.001)	336.833 (< 0.001)
	$c = 0, b$ free			21.735 (< 0.001)
	$b = 0, c$ free			1.751 (0.186)
$\lambda = 2, \nu = 2$	$b = c = 0$	306.023 (< 0.001)	230.990 (< 0.001)	306.114 (< 0.001)
	$c = 0, b$ free			0.091 (0.763)
	$b = 0, c$ free			75.124 (< 0.001)
$\lambda = \nu$	$b = c = 0$	307.949 (< 0.001)	280.136 (< 0.001)	310.810 (< 0.001)
	$c = 0, b$ free			2.861 (0.091)
	$b = 0, c$ free			30.674 (< 0.001)

Table 3 (continued)

Maintained hypothesis	H_0	H_A		
		$c = 0, b$ free	$b = 0, c$ free	b and c free
λ, v free	$b = c = 0$	309.179 (< 0.001)	275.798 (< 0.001)	310.654 (< 0.001)
	$c = 0, b$ free			1.475 (0.224)
	$b = 0, c$ free			34.857 (< 0.001)

The likelihood ratios test restrictions in the family of models given by $r_{t+1} = \mu + \gamma\sigma_t^2 + \sigma_t\varepsilon_{t+1}$ [Eq. (1) in text], $(\sigma_t^2 - 1)/\lambda = \omega + \alpha\sigma_{t-1}^2 f^\gamma(\varepsilon_t) + \beta((\sigma_{t-1}^2 - 1)/\lambda)$ [Eq. (7) in text], $f(\varepsilon_t) = \sqrt{a^2 + (\varepsilon_t - b)^2} - c(\varepsilon_t - b)$ [Eq. (6) in text], where r_{t+1} is the daily log return on a value-weighted equity index in excess of the risk-free interest rate, and σ_t is the conditional standard deviation of r_{t+1} . The numbers in parentheses are asymptotic probability values. All parameter estimates were obtained by maximum likelihood estimation under the assumption that $\varepsilon_t \sim N(0, 1)$, and a subset of the parameter estimates is given in Table 2.

The daily log excess stock returns, r_{t+1} , are measured as the log returns on the value-weighted index of NYSE stocks in excess of the log of the risk-free interest rate. The stock returns are a combination of the Schwert (1990) returns and the value-weighted index returns from the CRSP tapes. The riskless interest rates were obtained from the Fama/Bliss series, also on the CRSP tapes. The sample period – January 2, 1926 to December 31, 1990 – spans 17,486 observations.

for two-sided tests, allowing large positive shocks to have smaller or larger effects on future volatility than large negative shocks. Consequently, all of the symmetric GARCH models discussed earlier can be rejected in favor of asymmetric alternatives within the same class of models. In particular, the symmetric GARCH model of Bollerslev (1986) and the nonlinear GARCH model based on Bera and Higgins (1992) are rejected in favor of their asymmetric cousins.

Furthermore, several of the asymmetric models that have been proposed can be rejected in favor of the alternative that the asymmetry takes another form than the one specified by the model. For example, row three of the first panel shows that the standard EGARCH model of Nelson (1991) can be rejected in favor of an exponential GARCH model in which both a shift and a rotation of the news impact curve are present. In view of the sample size, however, this rejection is not very strong. On the other hand, row three of the third panel soundly rejects the Glosten, Jagannathan, and Runkle (1993) GARCH model with a rotated news impact curve in favor of a model with a rotated and shifted news impact curve. Since the test in the row above does not reject the hypothesis that the news impact curve is only shifted, the shift alone appears to be sufficient to capture the observed asymmetry within this class of models. Finally, the bottom row of panel four strongly rejects the rotated the news impact curve employed by Ding, Granger, and Engle (1993) in favor of a shifted one.

It has been argued previously that asymmetry is primarily a feature of volatility responses to large shocks. The results presented here suggest otherwise: Small shocks make asymmetric contributions to volatility. As discussed earlier, for small shocks, asymmetry is primarily driven by b , while the asymmetry due to c dominates for large shocks. In most of the models presented here, b is at least as important as c , suggesting that small shocks make significant asymmetric contributions to volatility.

The standard errors in Table 2 can be used to perform a few tests on λ and ν , but Table 4 gives a more complete set of likelihood ratio test statistics for these parameters. All models in this table permit both shifts and rotations of the news impact curve by freely estimating b and c . The appropriate parameter estimates are therefore found in the bottom rows of the appropriate panels in Table 2. The first column of the table lists the four null hypotheses of interest: The possibility that the data were generated by an EGARCH model, an AGARCH model, a GARCH model, or a nonlinear GARCH model. The second column tests the first and third null hypotheses against the alternative that the data were generated by a nonlinear GARCH model. As the test statistics show, both null hypotheses can easily be rejected. When λ and ν are restricted to be equal to one

Table 4
Likelihood ratio tests of functional form

H ₀	H _A	
	$\lambda = \nu$	λ, ν free
EGARCH		60.758
$\lambda = 0, \nu = 1$		(< 0.001)
AGARCH	75.089	81.443
$\lambda = 1, \nu = 1$	(< 0.001)	(< 0.001)
GARCH	21.563	27.918
$\lambda = 2, \nu = 2$	(< 0.001)	(< 0.001)
NARCH		6.355
$\lambda = \nu$		(0.012)

The likelihood ratios test restrictions in the family of models given by $r_{t+1} = \mu + \gamma\sigma_t^2 + \sigma_t\epsilon_{t+1}$ [Eq. (1) in text], $(\sigma_t^2 - 1)/\lambda = \omega + \alpha\sigma_{t-1}^2 f^*(\epsilon_t) + \beta((\sigma_{t-1}^2 - 1)/\lambda)$ [Eq. (7) in text], $f^*(\epsilon_t) = \sqrt{a^2 + (\epsilon_t - b)^2} - c(\epsilon_t - b)$ [Eq. (6) in text], where r_{t+1} is the daily log return on a value-weighted equity index in excess of the risk-free interest rate, and σ_t is the conditional standard deviation of r_{t+1} . The stock returns are a composite of the Schwert (1990) index returns and the CRSP index returns. The Fama/Bliss riskless rates were obtained from the CRSP bond files. The parameters b and c are freely estimated in all models. The numbers in parentheses are asymptotic probability values. All parameter estimates were obtained by maximum likelihood estimation and are given in Table 2. The sample runs from January 2, 1926 to December 31, 1990 and includes 17,486 observations.

another, they are clearly not equal to either one or two. Thus, they reject the AGARCH and GARCH specifications, even in their asymmetric forms. The EGARCH model is not nested within the alternative hypothesis, and no test statistic is reported. The third column, however, tests all four null hypotheses against the general alternative that the data were generated by a model in which λ and ν are different. Obviously, the second and third null hypotheses are also rejected by this even more general alternative. Additionally, the possibility that the data are generated by an exponential GARCH model is rejected in favor of the more general alternative. The χ^2_2 test statistic of 60.758 is far above all standard critical values. Lastly, the test of the restriction that $\lambda = \nu$ barely falls below the χ^2_1 1% critical value of 6.635. In view of the very large sample size, this rejection is rather weak; for many purposes it is probably defensible to describe the daily excess returns with models in which $\lambda = \nu$, but the exponent is either freely estimated, or at least between one and two.

Although the likelihood ratio tests indicate that the models differ from each other, the tests don't reveal how the models differ from each other. Fig. 3 plots the news impact curves based on the parameter estimates for the five fully asymmetric models in the bottom row of each panel of Table 2. All of the news impact curves in the figure assume that last period's conditional volatility, σ_{t-1} , was 0.01, which is very close to the sample mean. The volatilities were roughly scaled to annual percentage terms by a factor of 1,600. The figure shows the difference between the actual and expected conditional standard deviation for standardized shocks within four standard deviations of zero.⁶

The line corresponding to the AGARCH model is the only news impact curve with a negligible shift to the right and a minimum near zero. The remainder of the news impact curves are shifted to the right and have their minima between 0.322 and 0.488. The curves are close to one another near zero, but they diverge as the shocks grow in magnitude. For very large shocks, the GARCH model provides the largest innovation to volatility. Conversely, the AGARCH model provides the smallest responses to large shocks. Indeed, it was this very behavior that led Taylor (1986) and Schwert (1989) to suggest such a model. As discussed in Nelson and Foster (1994), the EGARCH model is almost as robust to large shocks as the AGARCH model, but nevertheless provides larger responses.

As the models with freely estimated exponents show, both types of responses are too extreme. The data prefer models in which large shocks increase volatility by more than they would in either the AGARCH or EGARCH models, but by less than they would in a GARCH model. The news impact curves for the nonlinear GARCH model ($\lambda = \nu$) and the model with freely estimated exponents

⁶This may appear to be an excessive range. Given the large sample size, however, one would expect to draw approximately eight standard normal innovations larger than three and a half, and almost fifty innovations larger than three.

λ and ν are virtually indistinguishable, but for large shocks, they clearly lie between the AGARCH and GARCH extremes. The fact that these two news impact curves are so close to one another is reflected in the weak rejection of the more restricted model by the likelihood ratio test.

Even though Fig. 3 covers a large range of shocks, it does not reveal the global behavior of the news impact curves. For extraordinarily large shocks, the EGARCH model provides the biggest innovation to volatility, while the fully

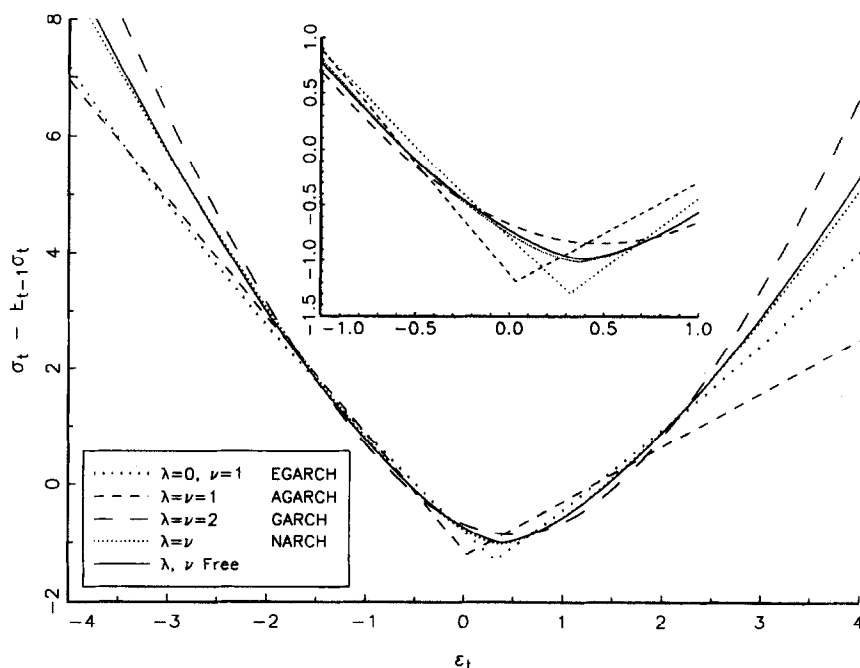


Fig. 3. News impact curves.

The figure shows the news impact curves for five different GARCH models. The news impact curves were estimated from the daily log return on a value-weighted equity index in excess of the risk-free interest rate. The stock returns are a composite of the Schwert (1990) index returns and the CRSP index returns. The Fama/Bliss riskless rates were obtained from the CRSP bond files. The entire sample period – January 2, 1926 to December 31, 1990 – spans 17,486 observations.

The conditional standard deviations, σ_t , of excess returns, r_{t+1} , were estimated from the family of models described by $r_{t+1} = \mu + \gamma \sigma_t^2 + \sigma_t \varepsilon_{t+1}$ [Eq. (1) in text], $(\sigma_t^2 - 1)/\lambda = \omega + \alpha \sigma_{t-1}^\lambda f'(\varepsilon_t) + \beta ((\sigma_{t-1}^2 - 1)/\lambda)$ [Eq. (7) in text], and $f(\varepsilon_t) = \sqrt{a^2 + (\varepsilon_t - b)^2} - c(\varepsilon_t - b)$ [Eq. (6) in text]. The appropriate restrictions on λ and ν were enforced, but b and c were freely estimated. The maximum likelihood parameter estimates for each of the models are given in the bottom rows of the five panels in Table 2. The innovations are conditional on $\sigma_{t-1} = 0.01$, the sample mean. When roughly converted to annual percentage terms with a scale factor of 1,600, this is equivalent to an annual volatility of 16%.

parameterized model runs second. For the parameter estimates in Table 2, however, the news impact curves of these two models only exceed the news impact curve of the GARCH model for shocks larger than fifteen or twenty standard deviations. The behavior of the news impact curves for such enormous shocks is merely a byproduct of the assumed functional form, not a meaningful reflection of the properties of the data – even in a data set of almost 18,000 observations.

To further explore the differences between the nested volatility models, Fig. 4 plots the volatility estimates from the EGARCH model of Nelson (1991), a model that freely estimates λ , v , b , and c , and the standard GARCH model of Bollerslev (1986). For the purposes of this figure, all quantities were averaged within calendar months due to the large volume of data. To roughly scale the volatility estimates to annual percentage terms, they were once again multiplied by 1,600.

The top row of plots in Fig. 4 shows the volatility estimates from each of the models. All three models clearly estimate the by-now-familiar patterns of volatility over the sample period. In annualized terms, volatility averages about 16% over the sample period, but was two to three times higher for an extended period in the late '20s and '30s, with occasional bursts of higher volatility thereafter. The plots confirm that the models respond to very large shocks according to the news impact curves shown in Fig. 3. Even when averaged over a whole month, very large shocks – like those of October 28, 1929 or October 19, 1987 – result in higher volatility in the GARCH model than in the freely estimated model or EGARCH model.

To highlight the differences between the models, the middle row of plots shows the differences between each pair of volatility estimates from the GARCH and EGARCH models, the GARCH and freely estimated models, and the freely estimated and EGARCH models. As the news impact curves of Fig. 3 suggest, the differences between the GARCH and EGARCH models are larger than those between the freely estimated model and either of the other two. During times of normal volatility, it is not unusual for these models to produce volatility estimates that differ by several percentage points. During times of high volatility, the differences occasionally reach double digits.

Furthermore, the bottom row of plots shows that the differences between any of these models, even when averaged over a whole month, can be considerable in relative terms. Most dramatically, the GARCH and EGARCH models can differ in their volatility estimates by monthly averages of more than 35%. In a day-by-day comparison, the GARCH model frequently estimates conditional variances that are more than twice as high as those from the EGARCH model.

4. Conclusion

This paper has developed a nested family of asymmetric GARCH models. The nesting of the models is accomplished by treating the variance equation as a law

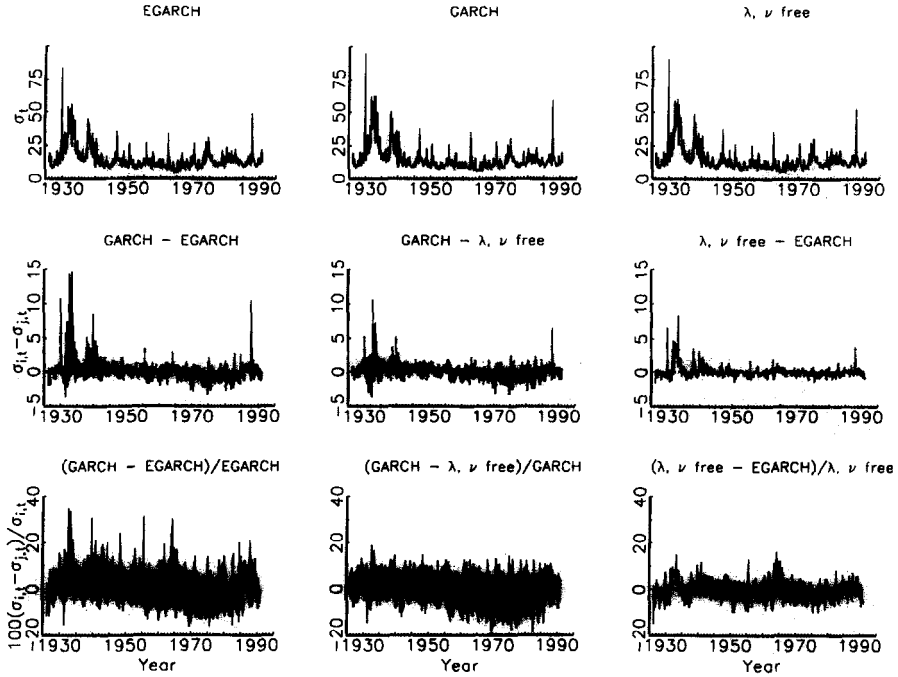


Fig. 4. Comparing conditional standard deviation estimates.

The plots compare estimates of the conditional standard deviation of the logarithms of daily returns on a value-weighted equity index in excess of the riskless interest rates. The stock returns are a combination of the Schwert (1990) returns and the daily CRSP index files. The interest rates were obtained from the Fama/Bliss files on the CRSP bond tapes. All plots average daily values within calendar months and the conditional standard deviations were roughly converted to annual percentage terms with a scale factor of 1,600.

The conditional standard deviations, $\sigma_{i,t}$, were estimated from the family of GARCH models given by $r_{t+1} = \mu + \gamma\sigma_t^2 + \sigma_t\varepsilon_{t+1}$ [Eq. (1) in text], $(\sigma_t^2 - 1)/\lambda = \omega + \alpha\sigma_{t-1}^\lambda f^\nu(\varepsilon_t) + \beta((\sigma_{t-1}^\lambda - 1)/\lambda)$ [Eq. (7) in text], and $f(\varepsilon_t) = \sqrt{a^2 + (\varepsilon_t - b)^2} - c(\varepsilon_t - b)$ [Eq. (6) in text], where r_{t+1} is the daily log return on a value-weighted equity index in excess of the risk-free interest rate. The GARCH estimates constrain $\lambda = \nu = 2$ and $b = c = 0$. The EGARCH estimates impose $\lambda = 0$, $\nu = 1$, and $b = 0$. The case described as λ, ν free does not restrict any of the parameters. All unrestricted parameters were estimated by maximum likelihood and the parameter estimates are given in Table 2.

The top row shows the estimates of the conditional standard deviation from each of the three models. The middle row shows the absolute difference between the estimates, while the bottom row shows the percentage difference between the models. In the bottom two rows, zero is marked by a horizontal dashed line.

The stock returns are a composite of the Schwert (1990) index returns and the CRSP index returns. The Fama/Bliss riskless rates were obtained from the CRSP bond files. The entire sample period – January 2, 1926 to December 31, 1990 – spans 17,486 observations.

of motion for the Box–Cox transformation of the conditional standard deviation. By properly selecting the shape of the Box–Cox transformation and the shape of the news impact curve, one can nest GARCH, absolute value GARCH, exponential GARCH, and other models.

The models in this family differ from each other in two respects. First, the members of the family employ different transformations of the conditional standard deviation. Standard GARCH describes the evolution of the conditional variance, AGARCH uses the conditional standard deviation itself, and EGARCH uses the natural logarithm of the conditional variance. Which of these transformations offers the most accurate description of actual variances can be tested based on a freely estimated exponent in the Box–Cox transformation. Tests based on daily U.S. stock returns reject all of the standard models in favor of a model that is, roughly speaking, a GARCH model for the conditional standard deviation. However, in this model the shifted absolute value of shocks raised to the power three halves drives the conditional standard deviation.

The second difference between the members of the family is their treatment of asymmetry. The well-known GARCH models either employ a shift or a rotation of the news impact curve, or ignore asymmetry altogether. This paper shows that for daily U.S. stock returns, a shift in the news impact curve dominates in most of the models that were estimated.

The GARCH family nests all of the popular GARCH models, but it also demonstrates that only a small number of possible models has been considered thus far. A striking feature of the news impact curves in Fig. 3 is that all of the models are similar. This is a reflection of the fact that all of the models are similarly constrained members of the same family of GARCH models. Indeed, the similarity of the news impact curves suggests that different families of models might provide fruitful avenues for future research. In particular, a class of models appears promising in which the transformation of the errors is more general than the simple asymmetric absolute value function raised to a power ν . Pagan and Schwert (1990) showed, however, that a purely sample-driven approach to this problem tends to lead to an overfitting of the data and poor out-of-sample performance. Consequently, theoretical insights into the relation between innovations and conditional volatility would be of considerable interest.

Appendix A: Positivity and covariance stationarity in the family of GARCH models

This appendix provides some regularity conditions for the family of GARCH models developed in the main text. Sufficient conditions for the positivity of the conditional variance are developed by analogy to the standard GARCH model,

and conditions for covariance stationarity of the shocks $\sigma_{t-1}\varepsilon_t$ are found by extension of Nelson (1990b).⁷

The family of GARCH models developed in the main text is given by

$$\frac{\sigma_t^\lambda - 1}{\lambda} = \omega' + \alpha \sigma_{t-1}^\lambda f^v(\varepsilon_t) + \beta \frac{\sigma_{t-1}^\lambda - 1}{\lambda}, \quad (\text{A.1})$$

$$f(\varepsilon_t) = |\varepsilon_t - b| - c(\varepsilon_t - b). \quad (\text{A.2})$$

For the purposes of this appendix, it is occasionally useful to rewrite Eq. (A.1) as

$$\sigma_t^\lambda = \omega + (\alpha \lambda f^v(\varepsilon_t) + \beta) \sigma_{t-1}^\lambda, \quad (\text{A.3})$$

when $\lambda \neq 0$. The intercept, ω , is defined as

$$\omega = \lambda \omega' - \beta + 1. \quad (\text{A.4})$$

A.1. Positivity

For $\lambda = 0$ or even integers $2/\lambda$, the conditional variance, σ_t^2 , is found by exponentiation, or by raising σ_t^λ to an even power. Either operation guarantees that σ_t^2 is nonnegative. Hence, positivity does not impose any restrictions on these models.

For the remaining cases, the intuition from the standard GARCH model can be used to find sufficient conditions for the positivity of σ_t^λ . Sufficient conditions for positivity of σ_t^λ , and therefore σ_t^2 , are that

$$\omega > 0, \quad (\text{A.5})$$

$$\alpha \geq 0, \quad (\text{A.6})$$

$$\beta \geq 0, \quad (\text{A.7})$$

$$|c| \leq 1 \quad \text{for } v \notin \{2, 4, 6, \dots\}. \quad (\text{A.8})$$

The last restriction is sufficient to guarantee that $f^v(\varepsilon_t)$ is nonnegative. The other restrictions ensure that σ_t^λ is the sum of positive elements and therefore positive itself. By analogy to the standard GARCH process of Bollerslev (1986), it is also possible to show that conditions (A.6) and (A.7) are not necessary for higher-order coefficients in higher-order GARCH models. Nelson and Cao (1992) have recently derived a set of weaker conditions for the standard GARCH process. Those conditions, as well as any future results, can be applied to this

⁷I gratefully acknowledge the extraordinarily helpful suggestions of Daniel Nelson, the referee. Much of the content of this appendix is a mere elaboration of these suggestions.

family of GARCH processes.⁸ To simplify the subsequent discussion, conditions (A.5)–(A.8) are assumed to hold unless stated otherwise.

A.2. Covariance stationarity

Sufficient conditions for the covariance stationarity of the family of GARCH(1, 1) processes can be derived by extension of Nelson's (1990b) work for the standard GARCH(1, 1) model. If one recursively substitutes for σ_{t-i}^λ in Eq. (A.3), one finds

$$\sigma_t^\lambda = \omega \left(1 + \sum_{k=0}^{\infty} \prod_{i=0}^k (\alpha \lambda f^v(\varepsilon_{t-i}) + \beta) \right). \quad (\text{A.9})$$

Theorems 1–3 from Nelson (1990b) now fit the GARCH process in Eq. (A.9). Sufficient conditions for the shocks $\eta_t = \sigma_{t-1} \varepsilon_t$ to be nondegenerate and strictly stationary-ergodic are that

$$\omega > 0 \quad \text{and} \quad E[\ln(\alpha \lambda f^v(\varepsilon_t) + \beta)] < 0.$$

Furthermore, from Nelson (1990b), we know that for any $p > 0$, $0 < E\sigma_t^{\lambda p} < \infty$ if $\omega > 0$ and $E[(\alpha \lambda f^v(\varepsilon_t) + \beta)^p] < 1$. Hence, the shocks $\sigma_{t-1} \varepsilon_t$ are covariance stationary and nondegenerate if

$$\text{var}(\varepsilon_t) < \infty, \quad (\text{A.10})$$

$$\omega > 0, \quad (\text{A.11})$$

and

$$E[(\alpha \lambda f^v(\varepsilon_t) + \beta)^{2/\lambda}] < 1. \quad (\text{A.12})$$

Under the assumption that $\varepsilon_t \sim N(0, 1)$, the first condition is satisfied by construction. The second and third conditions can either be enforced during estimation or verified afterwards.

For most of the models in this family, conditions (A.5)–(A.8) and (A.10)–(A.12) are both sufficient and necessary. The exceptions are formed by the aforementioned models that guarantee a positive conditional variance even when restrictions (A.5)–(A.8) are violated. If σ_t^λ is allowed to be negative but is restricted from changing sign, then the sign restriction on ω is reversed and all other conditions continue to apply. On the other hand, if the transformed volatility is

⁸Nelson and Cao (1992) provide necessary restrictions for the subset of all GARCH(p, q) models with $p = \{0, 1, 2\}$. The necessary restrictions in higher-order models are not yet known.

allowed to switch signs, for example by permitting $\beta < 0$, then necessary conditions for covariance stationarity are only known for a few special cases, such as the EGARCH model of Nelson (1991).

Under the assumption of normality, condition (A.12) can be evaluated analytically for some combinations of λ and ν .

A.2.1. Exponential GARCH ($\lambda = 0, \nu = 1$)

In the limit, as λ approaches zero and the variance equation converges to the class of EGARCH models, restriction (A.12) converges to

$$\beta < 1, \quad (\text{A.13})$$

which is the same as the size restriction given in Nelson (1991).

A.2.2. Absolute value GARCH ($\lambda = \nu = 1$)

For the class of AGARCH and TGARCH models with $\lambda = \nu = 1$, restriction (A.12) is equivalent to

$$\begin{aligned} &\alpha^2(1 + b^2)(1 + c^2) + \beta^2 + 2\alpha\beta bc + 4\alpha(\beta + \alpha bc)\phi(b) \\ &+ 2\alpha(\beta b + \alpha(1 + b^2)c)(2\Phi(b) - 1) < 1, \end{aligned} \quad (\text{A.14})$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the probability density and cumulative distribution function of the standard normal distribution, respectively. When $b = 0$, restriction (A.14) reduces to $\alpha^2(1 + c^2) + \beta^2 + 4\alpha\beta\phi(0) < 1$, which is equivalent to the restriction for the TGARCH(1, 1) model stated in Zakoian (1990).

A.2.3. GARCH ($\lambda = \nu = 2$)

For the class of standard GARCH models with $\lambda = \nu = 2$, restriction (A.12) can be evaluated as

$$\beta + 2\alpha\{(1 + b^2)(1 + c^2) + 2c[2\phi(b) + (1 + b^2)(2\Phi(b) - 1)]\} < 1. \quad (\text{A.15})$$

When $b = c = 0$, this restriction reduces to $\beta + 2\alpha < 1$, which is the restriction stated in Bollerslev (1986) when α is normalized by the factor 2.

Although similar calculations can be performed for other integer combinations of $2/\lambda$ and ν , other values generally make these calculations analytically intractable. Regardless of the parameter values and distributional assumptions, the conditions for covariance stationarity can always be verified numerically.

For higher-order models in this family that restrict the transformed volatility to be nonnegative, necessary and sufficient conditions for strict stationarity and ergodicity can be found by extension of Bougerol and Picard (1992). Unfortunately, the conditions for finite moments and hence covariance stationarity are not known for higher-order models.

Appendix B: Hyperbolic approximation of the absolute value function

Due to the nondifferentiability of the absolute value function, the model consisting of Eqs. (1), (5), and (6) in the main text generally has a nondifferentiable likelihood function. At best, this is a nuisance during numerical optimization of the likelihood function. Most of the commonly used numerical optimization algorithms make use of gradients and may not work properly on nondifferentiable functions. Although optimization routines that do not require gradients exist (e.g., Nelder and Mead, 1965; Powell, 1965), they do not compute an asymptotic covariance matrix for the parameter estimates, and thus make standard errors difficult to obtain. A simple but extremely effective remedy to this problem is to approximate the absolute value function by a hyperbola. That this is a natural approximation can be seen from the relation

$$\lim_{a \rightarrow 0} \sqrt{a^2 + \varepsilon_t^2} = |\varepsilon_t|. \quad (\text{B.1})$$

The function $\sqrt{a^2 + \varepsilon_t^2}$ is a rectangular hyperbola rotated counterclockwise by 45° . By decreasing the magnitude of a , the approximation can be improved to any desired accuracy while maintaining the differentiability of the likelihood function.

If analytical derivatives are used in the maximization of the likelihood function, this approximation is frequently unnecessary, since the derivatives are rarely evaluated exactly at the point of nondifferentiability. However, numerical gradients are usually computed by finite differences and benefit from this approximation, since they are more likely to straddle a point of nondifferentiability.

In practice, numerical optimization methods impose limits on the magnitude of a . For example, for very small values of a , numerical gradients computed by finite differences have the same problems with the hyperbolic approximation as they have with the absolute value function itself. Therefore, it is important to assess the accuracy of the approximation for moderate values of a . Since ε_t is i.i.d. with mean zero and unit variance, one can compare the moments of the hyperbolic approximation to the moments of the absolute value function at different values of a . In Table 5, such a comparison is provided for the first and second moments of normally distributed ε_t . The table compares the theoretical first and second moments of the absolute value of a standard normal ε_t to the corresponding numerically integrated moments of the hyperbolic approximation. The table shows that moderate values of a , 0.001 or 0.0001, for example, achieve excellent approximations of both the first and second moment. For these values of a , the relative error is on the order of a few parts per million or smaller, and does not affect any estimation results. Table 5 also shows that neither shifts nor rotations of the absolute value significantly affect the accuracy

Table 5
Accuracy of the hyperbolic approximation

a	b	c	$\mu_1^*/\mu_1 - 1$	$\mu_2^*/\mu_2 - 1$
0.1000	0.0	0.0	2×10^{-2}	1×10^{-2}
0.0100	0.0	0.0	3×10^{-4}	1×10^{-4}
0.0010	0.0	0.0	4×10^{-6}	1×10^{-6}
0.0001	0.0	0.0	5×10^{-8}	1×10^{-8}
0.0010	0.5	0.0	3×10^{-6}	8×10^{-7}
0.0010	0.0	0.5	4×10^{-6}	8×10^{-7}
0.0010	0.5	0.5	3×10^{-6}	5×10^{-7}

The ratios of the expectations, $\mu_i = E[|c - b| - c(c - b)]^i$ and $\mu_i^* = E[\sqrt{a^2 + (c - b)^2} - c(c - b)]^i$, given in the last two columns were computed as the ratios of the theoretical moments μ_i and the numerically integrated moments μ_i^* for $\varepsilon \sim N(0, 1)$.

The exact expectations are given by $\mu_1' = bc + 2\phi(b) + b(2\Phi(b) - 1)$ and $\mu_2' = (1 + b^2)(1 + c^2) + 2c[2b\phi(b) + (1 + b^2)(2\Phi(b) - 1)]$, where $\phi(\cdot)$ and $\Phi(\cdot)$ are the probability density and cumulative distribution function of the standard normal distribution respectively.

of the approximation. Hence, a value of 0.001 is used for a in all of the estimations. Other values of a were tried, but did not change the results unless they were almost two orders of magnitude larger than the value I use.

References

- Black, Fischer, 1976, Studies of stock price volatility changes, *Proceedings of the 1976 Meetings of the Business and Economics Statistics Section, American Statistical Association*, 177–181.
- Bollerslev, Tim, 1986, Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics* 31, 307–328.
- Bollerslev, Tim, 1987, A conditionally heteroskedastic time series model of security prices and rates of return data, *Review of Economics and Statistics* 59, 542–547.
- Bollerslev, Tim and Jeffrey M. Wooldridge, 1992, Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances, *Econometric Reviews* 11, 143–172.
- Bougerol, Phillipe and Nico Picard, 1992, Stationarity of GARCH processes and of some non-negative time series, *Journal of Econometrics* 52, 115–127.
- Box, G.E.P. and D.R. Cox, 1964, An analysis of transformations, *Journal of the Royal Statistical Society B* 26, 211–243.
- Christie, Andrew A., 1982, The stochastic behavior of common stock variances: Value, leverage, and interest rate effects, *Journal of Financial Economics* 10, 407–432.
- Davidian, M. and R.J. Carroll, 1987, Variance function estimation, *Journal of the American Statistical Association* 82, 1079–1091.
- Ding, Zhuanxin, Clive W.J. Granger, and Robert F. Engle, 1993, A long memory property of stock market returns and a new model, *Journal of Empirical Finance* 1, 83–106.
- Engle, Robert F., 1982, Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation, *Econometrica* 50, 987–1007.
- Engle, Robert F., 1990, Discussion: Stock market volatility and the crash of '87, *Review of Financial Studies* 3, 103–106.

- Engle, Robert F. and Tim Bollerslev, 1986, Modelling the persistence of conditional variances, *Econometric Reviews* 5, 1–50, 81–87.
- Engle, Robert F. and Victor K. Ng, 1993, Measuring and testing the impact of news on volatility, *Journal of Finance* 48, 1749–1778.
- Engle, Robert F., David M. Lilien, and Russell P. Robins, 1987, Estimating time varying risk premia in the term structure: The ARCH-M model, *Econometrica* 55, 391–407.
- French, Kenneth R., G. William Schwert, and Robert F. Stambaugh, 1987, Expected stock returns and volatility, *Journal of Financial Economics* 19, 3–29.
- Friedman, Benjamin M. and David I. Laibson, 1989, Economic implications of extraordinary movements in stock prices, *Brookings Papers on Economic Activity* 2, 137–189.
- Geweke, John, 1986, Comment, *Econometric Reviews* 5, 57–61.
- Glosten, Lawrence R., Ravi Jagannathan, and David E. Runkle, 1993, On the relation between the expected value and the volatility of the nominal excess return on stocks, *Journal of Finance* 48, 1779–1801.
- Higgins, Matt L. and Anil K. Bera, 1992, A class of nonlinear ARCH models, *International Economic Review* 33, 137–158.
- Hull, John and Alan White, 1987, The pricing of options with stochastic volatilities, *Journal of Finance* 42, 281–300.
- Johnson, Herb and David Shanno, 1987, Option pricing when the variance is changing, *Journal of Financial and Quantitative Analysis* 22, 143–151.
- Lumsdaine, Robin L., 1992, Asymptotic properties of the quasi-maximum likelihood estimator in GARCH(1, 1) and IGARCH(1, 1) models, Unpublished paper (Princeton University, Princeton, NJ).
- Mandelbrot, Benoit, 1963, The variation of certain speculative prices, *Journal of Business* 36, 394–419.
- Melino, Angelo and Stuart Turnbull, 1990, Pricing foreign currency options with stochastic volatility, *Journal of Econometrics* 45, 239–266.
- Nelder, J.A. and R. Mead, 1965, A simplex method for function minimization, *Computer Journal* 7, 308–313.
- Nelson, Daniel B., 1990a, ARCH models as diffusion approximations, *Journal of Econometrics* 45, 7–38.
- Nelson, Daniel B., 1990b, Stationarity and persistence in the GARCH(1, 1) model, *Econometric Theory* 6, 318–334.
- Nelson, Daniel B., 1991, Conditional heteroskedasticity in asset returns: A new approach, *Econometrica* 59, 347–370.
- Nelson, Daniel B. and Charles Q. Cao, 1992, Inequality constraints in the univariate GARCH model, *Journal of Business and Economic Statistics* 10, 229–235.
- Nelson, Daniel B. and Dean B. Foster, 1994, Asymptotic filtering theory for univariate ARCH models, *Econometrica* 62, 1–41.
- Pagan, Adrian R. and G. William Schwert, 1990, Alternative models for conditional stock volatility, *Journal of Econometrics* 45, 267–290.
- Pantula, Sastry G., 1986, Comment, *Econometric Reviews* 5, 71–73.
- Powell, M.J.D., 1965, An efficient method for finding the minimum of a function of several variables without calculating derivatives, *The Computer Journal* 7, 155–162.
- Rabemananjara, Roger and Jean-Michel Zakoian, 1993, Threshold ARCH models and asymmetries in volatility, *Journal of Applied Econometrics* 8, 31–49.
- Schwert, G. William, 1989, Why does stock market volatility change over time?, *Journal of Finance* 44, 1115–1153.
- Schwert, G. William, 1990, Indexes of United States stock prices from 1802 to 1987, *Journal of Business* 63, 399–426.
- Scott, Louis O., 1987, Option pricing when the variance changes randomly: Theory, estimation, and an application, *Journal of Financial and Quantitative Analysis*, 22, 419–438.

- Sentana, Enrique, 1991, Quadratic ARCH models: A potential reinterpretation of ARCH models as second-order Taylor approximations, Unpublished paper (London School of Economics and Political Science, London).
- Taylor, Stephen, 1986, *Modelling financial time series* (Wiley, New York, NY).
- White, Halbert, 1982, Maximum likelihood estimation of misspecified models, *Econometrica* 50, 1–25.
- Wiggins, James B., 1987, Option values under stochastic volatility: Theory and empirical estimates, *Journal of Financial Economics* 19, 351–372.
- Zakoian, Jean-Michel, 1991, Threshold heteroskedastic models, Unpublished paper (Institut National de la Statistique et des Etudes Economiques, Paris).