Lanne, Markku; Saikkonen, Pentti

Working Paper

Nonlinear GARCH models for highly persistent volatility

SFB 373 Discussion Paper, No. 2002,20

Provided in Cooperation with:
Collaborative Research Center 373: Quantification and Simulation of Economic Processes, Humboldt University Berlin

Suggested Citation: Lanne, Markku; Saikkonen, Pentti (2002) : Nonlinear GARCH models for highly persistent volatility, SFB 373 Discussion Paper, No. 2002,20, Humboldt University of Berlin, Interdisciplinary Research Project 373: Quantification and Simulation of Economic Processes, Berlin, https://nbn-resolving.de/urn:nbn:de:kobv:11-10048782

This Version is available at:
https://hdl.handle.net/10419/65348

Nonlinear GARCH Models for Highly Persistent Volatility*

by

Markku Lanne **
Department of Economics
University of Helsinki

Pentti Saikkonen***
Department of Statistics
University of Helsinki

Abstract

In this paper we study new nonlinear GARCH models mainly designed for time series with highly persistent volatility. For such series, conventional GARCH models have often proved unsatisfactory because they tend to exaggerate volatility persistence and exhibit poor forecasting ability. Our main emphasis is on models that are similar to previously introduced smooth transition GARCH models except for the novel feature that a lagged value of conditional variance is used as the transition variable. This choice of the transition variable corresponds to the idea that high persistence in conditional variance is related to relatively infrequent changes in regime. Using the theory of Markov chains we provide sufficient conditions for the stationarity and existence of moments of the considered smooth transition GARCH models and even some more general nonlinear GARCH models. Empirical applications to two exchange rate return series show that the new models can be superior to conventional GARCH models especially in longer term forecasting.

* Financial support from the Yrjö Jahnsson Foundation and the European Commission under the Training and Mobility of Researchers Programme is gratefully acknowledged. Part of this research was done while the first author was visiting the Institute of Statistics and Econometrics at the Humboldt University in Berlin.

** Research Unit on Economic Structures and Growth, Department of Economics, P.O. Box 54 (Unioninkatu 37), FIN-00014 University of Helsinki, FINLAND; e-mail: markku.lanne@helsinki.fi.

*** Department of Statistics, P.O. Box 54 (Unioninkatu 37), FIN-00014 University of Helsinki, FINLAND; e-mail: pentti.saikkonen@helsinki.fi

1 Introduction

In financial applications the linear GARCH model and some of its nonlinear extensions have been the most popular models for conditional variance. Recently, however, growing evidence has suggested that the typically observed very high persistence implied by these models does not characterize the behavior of exchange rates or stock returns. In particular, it has been demonstrated that GARCH models can exaggerate volatility persistence compared to the (true) volatility process perceived by the market (see Engle and Mustafa (1992) and Lamoureux and Lastrapes (1993)). This discrepancy seems to be especially pronounced after extreme shocks such as the October 1987 stock market crash. It has also been shown that observed high volatility persistence can be due to neglected nonlinearities such as level shifts (Lamoureux and Lastrapes (1990)) and that neglecting such potential nonlinearities can lead to poor forecasts (Hamilton and Susmel (1994)).

There are by now several alternative GARCH type models that attempt to take volatility persistence appropriately into account. These include the regime-switching ARCH models of Hamilton and Susmel (1994) and Cai (1994) and the fractionally integrated GARCH (FIGARCH) model of Baillie, Bollerslev and Mikkelsen (1996). A drawback of the former models is, however, that no GARCH terms are allowed within regimes, and, in practice, mere ARCH terms do not adequately, or at least not parsimoniously enough, capture the volatility persistence of many economic time series. As far as the FIGARCH model is concerned, it has the practical difficulty that in computations an infinite-order lag polynomial must be truncated at some arbitrary point and the choice of the truncation point can matter substantially when the model is applied (cf. Taylor (2000)). Besides, in order not to destroy important long-run dependencies the truncation point must be set at a relatively high value (typically 1000 is used), which necessitates the need of long time series in applications. Finally, as with the linear GARCH model, there is also evidence that the long memory in variance implied by the FIGARCH model can arise as a result of a structural change

(see Beine and Laurent (2001)).

In this paper we shall consider a nonlinear alternative of the conventional GARCH model. Our model is aimed at describing highly persistent volatility and it belongs to the family of smooth transition GARCH (STGARCH) models considered by Hagerud (1997), González-Rivera (1998), and Lundbergh and Teräsvirta (1998). However, unlike in these previous STGARCH models the transition variable of our model is a lagged value of the conditional variance and not a lagged value of the (squared) series. This appears to be a more suitable specification when the goal is to model series with persistent volatility. It will be seen that making only the level parameter in the conventional GARCH model change in this way can explain the observed high volatility persistence of some exchange rate return series. Our STGARCH model has features similar to the regime-switching GARCH models of Gray (1996) and Klaassen (in press): volatility persistence can depend on the level of conditional volatility in the previous period as well as on the size of the shock. However, compared to these regime-switching GARCH models, our model appears substantially more parsimonious.

In our empirical examples only a relatively simple STGARCH model turns out to be adequate. However, theoretical results will be obtained for very general nonlinear conditional variance models. A main limitation is that only a single ARCH term can be employed. From a practical point of view, this may not be a serious limitation because first order models are frequently found adequate in applications. The theoretical results of the paper give sufficient conditions for (a version of) geometric ergodicity and hence stationarity of both the volatility process and the observed process and they also establish existence of general integer order moments. These results allow for more general types of nonlinearity than the previous similar results of Carrasco and Chen (in press) and Ling and McAleer (in press, 2002). Proofs of these results make use of the stability theory developed for Markov chains. The aforementioned limitation of the considered general model is due to some difficulties with the application of this theory to models with more than a single ARCH term.

The rest of the paper proceeds as follows. Section 2 introduces the nonlinear

GARCH models to be studied in the paper and Section 3 presents the related assumptions. These assumptions are needed in Section 4 to obtain theoretical results of geometric ergodicity and existence of moments. Section 5 applies the new STGARCH models to two data sets, exchange return series of the German mark (DEM) and Japanese yen (JPY) against the U.S. dollar. The empirical results suggest that the new models manage to capture the volatility dynamics of these series better than linear GARCH models which particularly shows up as superior forecasting performance. Section 6 contains concluding remarks. Proofs of the main results are deferred to an appendix.

2 Nonlinear GARCH Models

A general model for the conditional variance of a real valued zero mean time series is

$$u_t = h_t^{1/2}\varepsilon_t, \quad t = 1, 2, ... \tag{1}$$

where h_t is a (measurable) function of u_s, $s < t$, and the (continuous) i.i.d.$(0, 1)$ random variables ε_t are independent of u_{t-1}, $u_{t-2}, \ldots$. These assumptions imply that u_t is a martingale difference sequence with conditional variance h_t. In the popular GARCH(1,1) model the conditional variance is specified as

$$h_t = \omega + \beta h_{t-1} + \alpha u_{t-1}^2, \quad t = 1, 2, ... \tag{2}$$

where the real valued parameters ω, β and α satisfy $\omega > 0$, $\beta \geq 0$ and $\alpha \geq 0$. This model can be extended to a GARCH(p,q) model by adding linear combinations of $h_{t-2}, \ldots, h_{t-p}$ and $u_{t-2}, \ldots, u_{t-q}$ to the right hand side of the equation. Conditions which ensure the strict and weak stationarity of the GARCH model can be found in Bollerslev (1986), Nelson (1990), and Bougerol and Picard (1992).[1] In the case of the simple GARCH(1,1) model (2) $\alpha + \beta < 1$ is sufficient for both strict and weak stationarity.

[1] Unless specified, the term stationary process will subsequently always refer to a strictly stationary process.

In practice the simple GARCH(1,1) model has been by far the most commonly used model for conditional variance. However, in many cases it has been found that estimates obtained for the parameters α and β are such that the stationarity condition $\alpha + \beta < 1$ is nearly violated (e.g., the sum of the estimates of α and β exceeds 0.98). As discussed in the introduction, models of this kind are often undesirable because they can exaggerate volatility persistence and, consequently, result in relatively poor volatility forecasts. Therefore, we shall consider nonlinear alternatives of the conventional GARCH model.

A special case of our general nonlinear alternative is given by

$$h_t = \omega + \delta_1 G_1(h_{t-1}) + \beta h_{t-1} + \alpha u_{t-1}^2, \quad t = 1, 2, ... \tag{3}$$

where ω, β and α are as in (2), $\delta_1 \geq 0$, and G_1: $\mathbb{R}_+ \to [0, 1]$ is an increasing function. In applications the function G_1 depends on parameters and it is supposed to be similar to the cumulative distribution function of a positive continuous random variable. A possible motivation of the function G_1 is to allow for a smooth shift in the parameter ω which determines the level of the conditional variance h_t. It will be seen in Section 5 that, at least in some cases, the observed nearly nonstationary behavior of the conventional GARCH(1,1) model (2) can be removed by a specification of this kind. In our empirical applications the function G_1 is chosen as the cumulative distribution function of a gamma distribution but any other similar function can also be considered (cf. Lundbergh and Teräsvirta (1998) who use a logistic function in a related situation).

Thus, the idea in our nonlinear model (3) is similar to that previously used in the STGARCH models considered by Hagerud (1997), González-Rivera (1998), and Lundbergh and Teräsvirta (1998). However, in all these previous studies the argument of the employed nonlinear function has been a lagged value of u_t^2 (or u_t) and not the conditional variance as in our model. In applications we have in mind it appears to be more useful to use $G_1(h_{t-1})$ in (3) instead of $G_1(u_{t-1}^2)$. Indeed, if $G_1(u_{t-1}^2)$ were used in (3) the level of the conditional variance would change whenever a small

(large) value of $|u_{t-1}|$ would be followed by a large (small) value of $|u_t|$. Since u_t is a martingale difference sequence this would imply very frequent changes in the level of the conditional variance. For instance, even if the process u_t evolves in a regime of high volatility it occasionally takes on small (absolute) values and, whenever this occurs, the conditional variance will drop to a lower regime. Because we are interested in modeling time series with highly persistent conditional variance such a behavior does not correspond to what we would expect to happen in reality. On the contrary, we would rather think that high persistence is related to rather infrequent changes in level. It may also be noted that the use of $G_1(h_{t-1})$ in (3) makes our model similar to autoregressive smooth transition models used for conditional expectation. In these models it is a lagged value of the series that is used to model the transition.

One might also be interested in an extension of model (3) in which the slope parameter β would shift in the same way as the level parameter ω. In other words, one might wish to extend model (3) by adding a term of the form $\delta G_1(h_{t-1})h_{t-1}$ to the right hand side. This would correspond to what has been considered in the aforementioned previous STGARCH models. We shall give theoretical results which also apply to such extensions although they are not needed in our empirical applications.

Motivated partly by potential further applications and partly by the desire to test the simple model (3) against more general alternatives we consider a general extension of the GARCH(p,1) model given by

$$h_t = g\left(h_{t-1}, \ldots, h_{t-p}\right) + f\left(u_{t-1}\right), \quad t = 1, 2, ... \tag{4}$$

where g: $\mathbb{R}_+^p \to \mathbb{R}_+$ and f: $\mathbb{R} \to [0, \infty)$ are nonlinear functions to be described in more detail in the next section. Here, as well as elsewhere, we use the notation $\mathbb{R}_+ = (0, \infty)$ and $\mathbb{R}_+^p = \mathbb{R}_+ \times \cdots \times \mathbb{R}_+$ (p copies).

Obviously, the previously discussed models (2) and (3) are special cases of (4). Both of them assume $f\left(u_{t-1}\right) = \alpha u_{t-1}^2$ which corresponds to the leading choice of the function f. Another specification of interest is

$$f\left(u_{t-1}\right) = \left(\alpha + \delta_2 G_2(u_{t-1})\right) u_{t-1}^2 \tag{5}$$

where $\alpha \geq 0$, $\delta_2 \geq 0$, and G_2: $\mathbb{R} \to [0,1]$ is an nondecreasing function which in applications may depend on parameters. Defining G_2 as the indicator function of the set $(0,\infty)$ yields the specification used in the GJR-GARCH model of Glosten, Jagannathan and Runkle (1993). A closely related alternative, suggested by Hagerud (1997), González-Rivera (1998), and Lundbergh and Teräsvirta (1998), is obtained by choosing G_2 as the logistic function. The motivation of these specifications is to allow for an asymmetric behavior in the conditional variance of previous GARCH models. The same idea may also be useful in the case of model (3).

Although our general model (4) covers a variety of interesting special cases it has the limitation that only a single lag of u_t is allowed. As far as the derivation of our theoretical results is concerned, it appears nontrivial to relax this assumption. To give an idea why this is the case and also to facilitate later developments, denote $X_{t-1} = [h_{t-1} \ \cdots \ h_{t-p}]'$ and combine equations (1) and (4) as

$$h_t = g\left(X_{t-1}\right) + f(h_{t-1}^{1/2}\varepsilon_{t-1}), \quad t = 1,2,\ldots . \tag{6}$$

Then define

$$F\left(X_{t-1}\right) = \left[F_1\left(X_{t-1}\right) \ \cdots \ F_p\left(X_{t-1}\right)\right]'$$

where $F_1\left(X_{t-1}\right) = g\left(X_{t-1}\right)$ and $F_i\left(X_{t-1}\right) = h_{t-i+1}$, $i = 2,...,p$. With this notation equations (4) and (6) can be expressed as

$$h_t = \iota' X_t \tag{7}$$

$$X_t = F\left(X_{t-1}\right) + H\left(X_{t-1},\varepsilon_{t-1}\right), \quad t = 1,2,..., \tag{8}$$

where $\iota' = [1\ 0\ \cdots\ 0]'$ $(p \times 1)$ and $H\left(X_{t-1},\varepsilon_{t-1}\right) = \left[f(h_{t-1}^{1/2}\varepsilon_{t-1})\ 0\ \cdots\ 0\right]'$ $(p \times 1)$. Since ε_{t-1} is independent of X_{t-1} it follows from (8) that X_t, $t = 1,2,...$ is a Markov chain.

The above discussion implies that the theory of Markov chains can be employed to prove the geometric ergodicity of the process X_t and hence that of h_t.[2] Once geometric

[2]We refer to Meyn and Tweedie (1993) for a comprehensive treatment of the needed Markov chain theory and to Chan (1990) and Tjøstheim (1990) for shorter reviews.

ergodicity has been established the existence of a stationary solution of equations (8) and (4) follow along with useful mixing results. From equation (1) one can then obtain similar results for the process u_t and for the joint process formed from u_t and h_t.

Adding more lags of u_t to model (4) complicates the application of the resulting Markov chain representation considerably. This particularly concerns proving the irreducibility of X_t which we need as an intermediate step to obtain geometric ergodicity. Since lags of u_t should be included in the model in the form $h_{t-i}^{1/2}\varepsilon_{t-i}$ the difficulty is similar to that encountered in bilinear models and generalized random coefficient models discussed by Pham (1986, p. 295). Carrasco and Chen (in press) avoid this difficulty because they use an approach which does not require proving irreducibility. However, this approach is not suitable in our case because the form of nonlinearity is considerably more general than permitted by Carrasco and Chen (in press). For these reasons we shall not try to extend the model by including more lags of u_t. A further reason for this is that a single lag of u_t has typically proved adequate in the applications of both linear and nonlinear GARCH models.

3 Assumptions

This section presents the assumptions needed to prove the subsequent theoretical results. Our first assumption is concerned with the innovation sequence ε_t and the function f in the general model (4).[3]

Assumption 1. (i) The i.i.d.$(0,1)$ random variables ε_t have a probability density function which is continuous and positive everywhere on $\mathbb{R}$. In addition, for some integer $k \geq 1$, $E\varepsilon_t^{2k} < \infty$.

(ii) The function f: $\mathbb{R} \to [0,\infty)$ is bounded on bounded subsets of $\mathbb{R}$, monotonically

[3]For ease of exposition, our assumptions rule out the possibility that f is constant. This is not restrictive, however, because it essentially means that we only rule out the trivial case where the conditional variance h_t is constant.

decreasing on $(-\infty, 0)$, and monotonically increasing on $(0, \infty)$. Moreover, $f(x)$ has a nonzero continuous derivative for $x \neq 0$.

(iii) There exists a nonconstant function b: $\mathbb{R} \to [0, \infty)$ such that $f(h_t^{1/2}\varepsilon_t) \leq h_t b(\varepsilon_t)$ and $b(x) \leq bx^2$ for some $b < \infty$.

The first assumption is standard in the statistical analysis of Markov chains and met in most cases where GARCH models are applied. From a practical point of view, the second assumption is not restrictive either and it could even be generalized considerably. In its present form Assumption 1(ii) implies that we can write

$$f(x) = f_1(x) + f_2(x)$$

where $f_1(x) = f(x)\mathbf{1}(x \leq 0)$ and $f_2(x) = f(x)\mathbf{1}(x > 0)$ with $\mathbf{1}(\cdot)$ the indicator function. The function $f_1(\cdot)$ is monotonically decreasing on $(-\infty, 0)$ and the function $f_2(\cdot)$ is monotonically increasing on $(0, \infty)$. The function $f(\cdot)$ may be nondifferentiable or even discontinuous at the origin. Assuming that the slope of the function $f(\cdot)$ changes at the origin instead of some other point is not essential. In a similar way we could allow for the possibility that the slope changes at any point. We have preferred to use the present form of Assumption 1(ii) because it is directly applicable to most of the models considered so far. The leading case $f(u_{t-1}) = \alpha u_{t-1}^2$ is a simple example. A slightly more complicated example is provided by the GJR-GARCH model of Glosten et al. (1993) (see the discussion after equation (5)).

By elementary probability calculus one can obtain the conditional density function of $f(h_{t-1}^{1/2}\varepsilon_{t-1})$ given $h_{t-1} = h$. If $\phi_\varepsilon(\cdot)$ denotes the common density function of the random variables ε_t the result can be expressed as

$$\psi(z; h) = h^{-1/2}\left|\frac{df_1^{-1}(z)}{dz}\right|\phi_\varepsilon\left(h^{-1/2}f_1^{-1}(z)\right) + h^{-1/2}\left|\frac{df_2^{-1}(z)}{dz}\right|\phi_\varepsilon\left(h^{-1/2}f_2^{-1}(z)\right) \quad (z > 0). \tag{9}$$

Here $f_1^{-1}(\cdot)$ is the inverse function of the restriction of $f_1(\cdot)$ to the interval $(-\infty, 0)$ with $f_1^{-1}(z) = 0$ for z not belonging to the range of $f_1(\cdot)$. The function $f_2^{-1}(\cdot)$ is defined similarly in terms of the function $f_2(\cdot)$ and the interval $(0, \infty)$. It is straight-

forward to check that $\psi(\cdot - g(x); x_1)$ is the conditional density function of h_t given $h_{t-i} = x_i$, $i = 1, ..., p$. This conditional density function is needed in our theoretical developments. However, these theoretical developments would allow us to assume more generally that, instead of two monotone functions, the function f can be expressed in terms of more than two such functions. For simplicity, we have preferred not to work with this extension explicitly.

Assumption 1(iii) restricts the increase of the function f to be at most quadratic. This requirement is met in most of the models considered so far. The formulation of this assumption may look somewhat peculiar but it has real advantages over its simpler alternative $f(x) \leq bx^2$. To illustrate this point, consider the GJR-GARCH model of Glosten et al. (1993). As pointed out after equation (5), this specification can be obtained by defining $G_2(u_{t-1})$ therein as $G_2(u_{t-1}) = \mathbf{1}(u_{t-1} > 0)$. From equation (5) it then follows that Assumption 1(iii) could be used with $b(x) = (\alpha + \delta_2)x^2$. This essentially means that one ignores the first inequality in Assumption 1(iii) and uses only the condition $f(x) \leq bx^2$. However, since $\mathbf{1}(u_{t-1} > 0) = \mathbf{1}(\varepsilon_{t-1} > 0)$ we can write

$$f(h_t^{1/2}\varepsilon_t) = h_t(\alpha + \delta_2 \mathbf{1}(\varepsilon_{t-1} > 0))\varepsilon_t^2. \tag{10}$$

This shows that Assumption 1(iii) can also be used with $b(x) = (\alpha + \delta_2 \mathbf{1}(x > 0))x^2$. It will be seen later that to obtain geometric ergodicity we need a bound for $Eb(\varepsilon_t)$ and it will be advantageous to have this bound as small as possible. If one uses Assumption 1(iii) with $b(x) = (\alpha + \delta_2 \mathbf{1}(x > 0))x^2$ and imposes the commonly used additional assumption that the innovations have a symmetric distribution one obtains $Eb(\varepsilon_t) = \alpha + \delta_2/2$. This is smaller than $Eb(\varepsilon_t) = \alpha + \delta_2$ obtained when Assumption 1(iii) is used with $b(x) = (\alpha + \delta_2)x^2$.

The next assumption imposes restrictions on the function g in (4). When applied to a vector, the symbol $|\cdot|$ will signify the norm defined by $|v| = \sum_{i=1}^{p} |v_i|$.

Assumption 2. (i) The function g: $\mathbb{R}_+^p \to \mathbb{R}_+^p$ is measurable and bounded on bounded subsets of $\mathbb{R}_+^p$. Moreover, for some $\eta > 0$, $\inf_{x \in \mathbb{R}_+^p} g(x) = \eta$.

(ii) There exists a vector $a = [a_1 \ \cdots \ a_p]'$ with $a_i \geq 0$, $i = 1, \ldots, p$, and such that $g(x) \leq a'x + o(|x|)$ as $|x| \to \infty$.

The first assumption is very mild. The requirement that the function g has a positive lower bound is assumed for technical reasons to facilitate mathematical derivations. As our previous examples demonstrate, this assumption is met in most of the models considered so far. Conditions of the type required for the function g in Assumption 2(ii) have also been used previously when Markov chain theory is applied to obtain stability results for time series models (see Masry and Tjøstheim (1995), Lu (1996, 1998), and Lu and Jiang (2001)). As with Assumption 1(iii), this assumption is also satisfied in the standard GARCH(p,1) model as well as in (3) whose extension will be discussed in the next section.

The last condition in assumption 2(i) implies that $h_t \geq \eta$ for all $t \geq 1$. In what follows, we shall therefore assume that the state space of the Markov chain X_t is $\mathcal{X} = [\eta, \infty) \times \cdots \times [\eta, \infty)$.

4 Geometric Ergodicity

Given the representation (8) and Assumptions 1 and 2, we can show the geometric ergodicity of the process h_t defined in (6). Instead of the usual version of geometric ergodicity we shall employ the so-called V-geometric ergodicity of a Markov chain (see Meyn and Tweedie (1993, p. 356)). Here V signifies a real valued function defined on the state space of the considered Markov chain and such that $V(\cdot) \geq 1$. For such a function V, the Markov chain X_t is said to be V-geometrically ergodic if there exists a probability measure π on the Borel sets of $\mathcal{X}$ and a constant $\varrho > 1$ such that

$$\sum_{t=1}^{\infty} \varrho^t \sup_{q:|q|\leq V} \left| E\left(q\left(X_t\right) \ \mid X_0 = x\right) - \int_{\mathcal{X}} \pi\left(dy\right) q\left(y\right) \right| < \infty \quad \text{for all } x \in \mathcal{X}. \tag{11}$$

The definition also assumes that the function V is integrable with respect to the probability measure π. The weakest form of this definition results when $V(\cdot) \equiv 1$. Then the Markov chain X_t is said to be geometrically ergodic. Geometric ergodicity

entails that the t-step transition probability measure $P^t(x,\cdot)$ defined on the Borel sets of $\mathcal{X}$ by $P^t(x,A) = P(X_t \in A \mid X_0 = x)$ converges at a geometric rate and for all $x \in \mathcal{X}$ to the probability measure $\pi(\cdot)$ with respect to the total variation norm. Note that the conditional expectation in (11) is defined in terms of the one-step transition probability measure $P^1(x,\cdot) = P(x,\cdot)$.

It is straightforward to show that geometric ergodicity implies stationarity of the process X_t if the distribution of the initial value X_0 is defined by the probability measure π (see Meyn and Tweedie (1993, p. 230-231)). Therefore, π is often referred to as the stationary probability measure of X_t. Of course, the stationarity of the process X_t implies the stationarity of h_t. A convenient feature of V-geometric ergodicity is that it immediately shows existence of moments. Specifically, when the Markov chain X_t is initialized from the stationary distribution, V-geometric ergodicity implies that the expectation of $q(X_t)$ exists for all q such that $|q(\cdot)| \leq V(\cdot)$.

In the following theorem we show that the Markov chain X_t is $|x|^k$-geometrically ergodic where the integer $k \geq 1$ is such that $E\varepsilon_t^{2k} \leq \infty$ (see Assumption 1(i)). To be able to formulate this theorem, we define the companion matrix

$$A = \begin{bmatrix} a_1 & a_2 & \cdots & a_{p-1} & a_p \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \quad (p \times p) \tag{12}$$

and the matrix

$$B(\varepsilon_t) = b(\varepsilon_t)\, \iota\iota' \quad (p \times p). \tag{13}$$

Here the notation is as in Assumption 1(iii) and (8) so that the matrix $B(\varepsilon_t)$ has zero elements except for $b(\varepsilon_t)$ in the (1,1) position. We use $\rho(\cdot)$ to signify the largest absolute eigenvalue of the indicated matrix or the so-called spectral radius. Furthermore, for any square matrix M, $M^{\otimes k} = M \otimes \cdots \otimes M$ denotes Kronecker's product taken k times. Now we can prove the following theorem.

Theorem 1 *Consider the Markov chain X_t in (8) and suppose that Assumptions 1 and 2 hold. Then, if $\rho\left(E\left(A+B(\varepsilon_t)\right)^{\otimes k}\right)<1$ the following results hold.*
(a) X_t is $|x|^k$-geometrically ergodic and the same is true for the joint process (u_t,h_t). Hence, $Eh_t^k<\infty$ and $Eu_t^{2k}<\infty$.
(b) If the distribution of the initial values has finite moments of order k then the process (u_t,h_t) is strong mixing with geometrically decaying mixing numbers.
(c) If the initial values have the stationary distribution the process (u_t,h_t) is β-mixing (absolutely regular) with geometrically decaying mixing numbers.

As our previous discussion makes clear, the first result of Theorem 1 implies that there exists a choice of initial values such that the process (u_t,h_t) is stationary and has finite moments of order k. A proof for the result $Eu_t^{2k}<\infty$ is obtained from this and equation (1). Parts (b) and (c) of Theorem 1 are useful because they make it possible to apply conventional limit theorems needed in the development of asymptotic estimation and testing procedures. A definition of the concept of β-mixing can be found in Doukham (1994, Section 1.1) where it is also shown that β-mixing implies strong mixing with mixing numbers converging to zero at least at the same rate. It would be possible to relax the assumption imposed on initial values in part (c) of Theorem 1 but something stronger than assumed in part (b) is needed (see Doukham (1994, p. 89 and 92)).

The eigenvalue condition used in Theorem 1 is simple to check in the leading cases where $k=1,2$ and $b\left(\varepsilon_t\right)=b\varepsilon_t^2$. For $k=1$, this condition becomes $\rho\left(A+b\iota\iota'\right)<1$, which is equivalent to $\sum_{j=1}^{p}a_j+b<1$ because $a_j\geq 0$ $(j=1,...,p)$ and $b>0$. If $k=2$ the condition also involves the fourth moment of ε_t. Then, if $\varepsilon_t\sim N\left(0,1\right)$ and $p=1$, the condition $\rho\left(A+b\iota\iota'\right)<1$ reduces to $a_1^2+2a_1b+3b^2<1$.

Theorem 1 can be specialized to various special cases. In our empirical applications we are interested in models which are special cases of

$$h_t=\omega+\beta_1h_{t-1}+\cdots+\beta_ph_{t-p}+\delta_1G_1(h_{t-d})+\left(\alpha+\delta_2G_2(u_{t-1})\right)u_{t-1}^2,\quad t=1,2,... \quad (14)$$

where $1\leq d\leq p$. This model extends the conventional GARCH(p,1) model by

simultaneously allowing for nonlinearities of the form discussed in the contexts of the equations (3) and (5). From Theorem 1 we can easily obtain the following result which applies to the GJR-GARCH model of Glosten et al. (1993) and the analogous alternatives proposed by Hagerud (1997), González-Rivera (1998), and Lundbergh and Teräsvirta (1998).

Corollary 2 *Let h_t be generated by (14) where u_t is as in (1), $\omega > 0$, $\beta_i \geq 0$ $(i = 1, ..., p)$, $\delta_i \geq 0$ $(i = 1, 2)$ and $\alpha \geq 0$. Assume that G_1: $\mathbb{R}_+ \to \mathbb{R}_+$ is a measurable bounded function and G_2: $\mathbb{R} \to [0, 1]$ is such that Assumption 1(ii) holds with $f(x) = (\alpha + \delta_2 G_2(x))\, x^2$. Assume further that $\rho\left(E\left(A + B(\varepsilon_t)\right)^{\otimes k}\right) < 1$ with the matrices A and $B(\varepsilon_t)$ defined by choosing $a_i = \beta_i$ $(i = 1, ..., p)$ and $b(\varepsilon_t) = (\alpha + \delta_2)\varepsilon_t^2$. Then, if ε_t in (1) satisfies Assumption 1(i) the results of Theorem 1 hold.*

Corollary 2 can be proved by verifying the assumptions of Theorem 1. Assumptions 1(i) and (ii) hold trivially whereas, by the definitions of the functions G_2 and f, Assumption 1(iii) holds if we take $b(x) = (\alpha + \delta_2)x^2$ and $b = \alpha + \delta_2$. As for Assumption 2, its first part is immediate and the validity of the second part follows from the choice $a_i = \beta_i$ $(i = 1, ..., p)$ and the boundedness of the function G_1.

When $\delta_2 = 0$ and $k = 1$ the condition imposed on the parameters α and β_i $(i = 1, ..., p)$ in Corollary 2 agrees with previous stationarity conditions obtained for the standard GARCH(p,q) model with $q = 1$ (see Bougerol and Picard (1992) for strict stationarity and Bollerslev (1986) for weak stationarity). Proposition 12 of Carrasco and Chen (in press) proves a result similar to Corollary 2 for power GARCH(p,q) models whereas Ling and McAleer (in press) give conditions for the existence of moments in a vector ARMA-GARCH model. Corollary 2 provides a (partial) extension of these previous results by including the bounded nonlinear terms $G_1(h_{t-1})$ and $G_2(u_{t-1})$ in the model.

Although Corollary 2 applies to the GJR-GARCH model a better result can be obtained in this case, as indicated after Assumption 1. Specifically, we have the following result.

Corollary 3 *If* $G_2(x) = \mathbf{1}(x > 0)$ *in Corollary 2 the stated result holds with* $b(\varepsilon_t)$ *defined as* $b(\varepsilon_t) = (\alpha + \delta_2 \mathbf{1}(\varepsilon_t > 0))\varepsilon_t^2$.

As the discussions after equation (10) and Theorem 1 show, if the innovations have a symmetric distribution the application of Corollary 3 in the case $k = 1$ gives the condition $\sum_{j=1}^{p} \beta_j + \alpha + \delta_2/2 < 1$ instead of $\sum_{j=1}^{p} \beta_j + \alpha + \delta_2 < 1$ obtained from Corollary 2. When $\varepsilon_t \sim N(0,1)$, $k = 2$ and $p = 1$ we get the condition $\beta_1^2 + 2\alpha\beta + 3\alpha^2 + \beta_1\delta_2 + 3\alpha\delta_2 + 3\delta_2/2 < 1$ previously obtained in Example 2.1(ii) of Ling and McAleer (2002) for the conventional GJR-GARCH(1,1) model.

Our final application of Theorem 1 is concerned with the first order model

$$h_t = \omega + \delta_1 G_1(h_{t-1}) + (\beta + \delta_3 G_1(h_{t-1}))\, h_{t-1} + (\alpha + \delta_2 G_2(u_{t-1}))\, u_{t-1}^2, \quad t = 1, 2, \ldots . \tag{15}$$

Thus, in this case also the coefficient of h_{t-1} is allowed to change. However, unlike in Corollaries 2 and 3, we shall now assume that G_1 is similar to the cumulative distribution function of a positive continuous random variable. From Theorem 1 the following result can then be obtained.

Corollary 4 *Let* h_t *be generated by (15) where* u_t *is as in (1). Assume that* G_1*:* $\mathbb{R}_+ \to [0,1]$ *is an nondecreasing function such that* $G_1(x) \to 1$ *as* $x \longrightarrow \infty$. *Assume further that* $\beta + \delta_3 \geq 0$ *and that* ω, δ_i $(i = 1, 2)$, α *and* G_2 *are as in Corollary 2. Then, if* ε_t *in (1) satisfies Assumption 1(i) and* $E(\beta + \delta_3 + b(\varepsilon_t))^k < 1$, *where* $b(\varepsilon_t) = (\alpha + \delta_2)\varepsilon_t^2$, *the results of Theorem 1 hold. Moreover, if* $G_2(x) = \mathbf{1}(x > 0)$ *the same conclusion obtains if* $b(\varepsilon_t)$ *is defined as* $b(\varepsilon_t) = (\alpha + \delta_2 \mathbf{1}(\varepsilon_t > 0))\varepsilon_t^2$.

According to what was said to justify Corollary 2, we only need to verify Assumption 2 whose first part is obviously satisfied. Moreover, now $p = 1$ and

$$g(x) = \omega + \delta_1 G_1(x) + (\beta + \delta_3 G_1(x))\, x = (\beta + \delta_3)\, x + o(x) \text{ as } |x| \longrightarrow \infty.$$

Here the latter equality follows from the assumptions made of the function G_1. Thus, Assumption 2(ii) holds with $a_1 = \beta + \delta_3$ and the stated result is obtained because in the case of Corollary 4 the assumed moment conditions are equivalent to

$\rho\left(E\left(A+B(\varepsilon_t)\right)^{\otimes k}\right)<1$ used in Theorem 1. If $k=1$ the first moment condition in Corollary 4 reduces to $\beta+\delta_3+\alpha+\delta_2<1$ and, if the distribution of ε_t is symmetric, the second one becomes $\beta+\delta_3+\alpha+\delta_2/2<1$. Setting $\delta_3=0$ yields the conditions obtained from Corollaries 2 and 3 in the case $p=1$ and $k=1$. Note that these conditions require $\beta<1$ whereas in Corollary 4 we may have $\beta>1$ because $\delta_3<0$ is possible.

It would be of interest to consider extensions of models (14) and (15) and make the coefficients of $h_{t-1},...,h_{t-p}$ in the former depend on h_{t-d} $(1\leq d\leq p)$ in the same way as the corresponding coefficient depends on h_{t-1} in (15). However, such extensions are not easy to handle with Theorem 1 and therefore we leave this topic for future research.

5 Empirical Illustration

To illustrate the properties of our nonlinear STGARCH models applications to two exchange rate return series are presented. The model we consider is given by (1) and (3) with $G_1(h_{t-1})$ the cumulative distribution function of the gamma distribution, that is,

$$G_1(h_{t-1})=\int_0^{h_{t-1}}\frac{(\gamma/c)^\gamma}{\Gamma(\gamma)}s^{\gamma-1}e^{-\gamma s/c}ds$$

where $\Gamma(\cdot)$ is the gamma function. In practice, estimation is greatly facilitated by using the standard form of the gamma distribution in which the restriction $\gamma=c$ is employed. According to our experience, relaxing this restriction has a negligible effect on the value of the likelihood function and estimates of the other parameters of the model. Therefore, this restriction is assumed in the empirical models. The model obtained from (1) and (3) with this restriction will be referred to as the STGARCH(1,1) model.[4] Results concerning the corresponding linear GARCH model defined by (1) and (2) are provided for comparison.

[4]We also considered a model where the term $\delta_1 G_1(h_{t-1})$ was replaced by $\delta G_1(h_{t-1})h_{t-1}$. For these series, however, this model turned out to be dominated by the linear GARCH(1,1) model.

Once a distribution for ε_t in (1) has been specified the parameters of the STGARCH model described above can be straightforwardly estimated by the method of maximum likelihood. As many financial time series, including exchange rate returns, are leptokurtic, quasi maximum likelihood estimation based on the conditional normal distribution may not be optimal, and instead some leptokurtic distribution such as the t distribution may be a more appropriate choice. Asymptotic properties of such quasi maximum likelihood estimators are unknown. However, assuming that the parameters are located in the stationary region given in Theorem 1 or Corollary 3 it is reasonable to expect that standard asymptotic results of statistical inference apply.

To check the adequacy of the estimated models analogues of standard diagnostic tests developed for linear GARCH models will be employed to evaluate the STGARCH models. In particular, the hypotheses of no remaining ARCH, no unmodeled autocorrelation, symmetry and parameter constancy are of interest. The asymmetry we consider is of the GJR-GARCH type (Glosten et al. (1993)) discussed in Section 2. When applied to the estimated linear GARCH model, especially the test for parameter constancy is likely to have power against the STGARCH model. Recently, Lundbergh and Teräsvirta (in press) introduced robust versions of LM tests for some of these hypotheses and showed that they are superior when the errors are nonnormal, as is likely to be the case with most financial time series. Therefore, these tests and their analogues for the STGARCH model will be employed.

The data set consists of the daily spot rates of the DEM and JPY against the U.S. dollar covering the period from the beginning of 1980 until the end of July 1998 (4773 observations), obtained from Datastream[5]. The returns are computed as logarithmic differences multiplied by 100. Daily exchange rate return series can be, in general, characterized by strong volatility persistence. Indeed, the sum of the estimated α and β coefficients of a standard GARCH(1,1) model fitted to the DEM and JPY return series equal 0.984 and 0.988, respectively, potentially hinting at an integrated GARCH model. The series are depicted in Figure 1. They have both tranquil and

[5]The dataset was downloaded from the internet web site of Franses and van Dijk (2000).

volatile periods with the JPY return series having an exceptionally volatile period at the end of the sample. The estimated kurtosis of the DEM return series is 4.921 and that of the JPY return series 6.897, each clearly exceeding the value 3 implied by normality.

The estimation results of the nonlinear STGARCH(1,1) and GARCH(1,1) models are presented in Table 1. The models for the mean-adjusted series were estimated by the method of maximum likelihood assuming t distributed errors to take the excess kurtosis into account[6]. The estimates of the degrees of freedom parameter ν are precisely estimated and indicate deviations from conditional normality. According to asymptotic standard errors all the other parameters are also very accurately estimated. As was mentioned above, the sums of α and β in the linear GARCH models are estimated close to unity. For the nonlinear STGARCH models, the parameter estimate of β was not significantly different from zero, and hence only models assuming $\beta = 0$ are considered (the p values of the likelihood ratio test for the null hypothesis $\beta = 0$ were 0.159 and 0.095 for the DEM and JPY return series, respectively). When both h_{t-1} and $G_1(h_{t-1})$ were included in the model, both of the parameters β and δ_1 turned out to be imprecisely estimated, indicating that to some extent they are close substitutes. As will be discussed below, this seems to be true only at low levels of conditional variance, though. The estimated α coefficients obviously satisfy the condition in Theorem 1 or Corollary 3, guaranteeing stability of the models.

Table 1 also reports the marginal significance levels of the diagnostic tests mentioned above. The null hypotheses of no remaining (first order) autocorrelation and ARCH up to order 10 cannot be rejected at the 5% level for either model. For the DEM return series both models also seem to be able to capture potential asymmetries, while there is evidence of unmodeled asymmetry in the linear GARCH model for the JPY return series. For the corresponding nonlinear STGARCH model the null of no asymmetry, in contrast, cannot be rejected. Although the news impact curve (see below) of the STGARCH model is symmetric, the model can to some extent

[6]The models were also estimated assuming normality, and the results were qualitatively similar.

alleviate asymmetric effects if, for instance, large negative shocks typically coincide with high conditional volatility as seems to be case for the JPY return series. The model then moderates the effect of these shocks compared to the corresponding linear GARCH model. Results of the parameter constancy tests are not alarming for any of the models although slight evidence of parameter nonconstancy may be seen for the DEM series.

The properties of the nonlinear STGARCH models can best be described by considering their implied news impact curves (Pagan and Schwert, 1990) which show the relationship between the current shock u_t and conditional volatility in the next period, h_{t+1}, keeping all other information constant. For the linear GARCH(1,1) model the news impact curve is thus defined as

$$NIC(u_t|h_t = h) = \omega + \alpha u_t^2 + \beta h$$

which is a parabola centered at $u_t = 0$ indicating that the impact of a shock on next period's conditional volatility is increasing in its absolute value. The value of the conditional volatility h only moves the curve vertically and the moves are proportional to h. For the STGARCH(1,1) model, on the other hand, the NIC is given by the following equation

$$NIC(u_t|h_t = h) = \omega + \alpha u_t^2 + \delta_1 G(h;\gamma)$$

which is also symmetric around zero. Here the value of h also moves the curve vertically but the size of these moves depends on h such that the effect of h on the NIC diminishes with increasing conditional volatility. This is depicted in Figure 2 which displays the NICs of the GARCH(1,1) and STGARCH(1,1) models as functions of h at $u_t = 1$. For the STGARCH models the effect of a shock grows rapidly with h at low levels of conditional volatility, but when the conditional volatility is initially high this growth levels off, hence moderating the effect of shocks at high conditional volatility levels. This is in contrast with the linear GARCH(1,1) model where the corresponding curves are straight lines, indicating proportional dependence of the impact of a shock

on the size of initial conditional volatility. The curves for the GARCH models also lie above those for the STGARCH models at all levels of conditional variance with the difference increasing as a function of h. Although over 95% of the values of the conditional variance implied by all the models fall below 1.0, where this difference does not seem to be large, for a substantial amount of observations the effects implied by the GARCH models deviate considerably from those implied by the STGARCH models (the maximum values of conditional variance implied by the GARCH models are 2.0 and 4.2 for the DEM and JPY returns, respectively).

Further insight into the differences between the estimated models can be gained by examining their dynamic properties. To this end we have computed the cumulative impulse response functions implied by the estimated models. The cumulative impulse response functions give the cumulative effect of a shock in period t, u_t, on the conditional volatility s periods ahead conditional on the size of the shock, λ, and the value of h_t. For GARCH(1,1) models the impulse response function can be computed analytically but for STGARCH models the following simulation technique is employed. Given initial values $h_t = h$ and $u_t = \lambda$, the model is simulated s periods ahead to obtain N realizations, and these are averaged to approximate the so-called conditional volatility profile (see Gallant, Rossi and Tauchen (1993))

$$E(h_{t+s}|h_t = h, u_t = \lambda), \quad s = 1, 2, ...$$

The cumulative impulse response function $IRF\,(h, s, \lambda)$ is then obtained as the difference between this profile and the one assuming a zero initial shock,

$$IRF\,(h, s, \lambda) = E(h_{t+s}|h_t = h, u_t = \lambda) - E(h_{t+s}|h_t = h, u_t = 0).$$

The shape of the cumulative impulse response functions of the GARCH model does not depend on λ or h whereas that of the STGARCH model can depend on both.

Cumulative impulse response functions for the STGARCH(1,1) models estimated for the DEM and JPY returns are depicted in the top and middle panels of Figures 3 and 4 for $\lambda = 1$ and 2, respectively. Three different values of h are considered,

0.1, 0.5 and 2.0. The middle value is rather close to the unconditional variance of both return series, while 0.1 represents a "low" and 2.0 "high" conditional volatility. When the shock is small ($\lambda = 1$) the decay of the cumulative IRF is slowest near the unconditional volatility and fastest in the case of high initial conditional volatility. For the JPY return series the ranking of the cumulative IRFs is similar when $\lambda = 2$, while the IRFs implied by the DEM return model decay slowest when $h = 0.1$, and the difference between the $h = 0.1$ and $h = 0.5$ cases is minor. Most importantly, the decay is always fastest when the initial conditional volatility is high, indicating the transient nature of very volatile periods, i.e., when the volatility is already high, the effect of a shock is only temporary.

The bottom panels of Figures 3 and 4 depict the cumulative impulse response functions for the GARCH(1,1) models assuming $\lambda = 1$. For both return series the decay is much slower for the GARCH(1,1) models. The contrast to the STGARCH models is consistent with the finding that linear GARCH models tend to exaggerate the persistence of volatility (see e.g. Lamoureux and Lastrapes (1993) for evidence on stock market returns based on comparison with volatility implicit in option prices). Moreover, the GARCH model may also be too restrictive if there indeed are nonlinearities such that the market perceives shocks in the high volatility regime as having different persistence properties. Klaassen's (in press) regime-switching GARCH models for exchange rate returns and Engle and Mustafa's (1992) results indicating that linear GARCH models exaggerate volatility persistence after the 1987 stock market crash, lend support to this kind of behavior.

To further compare the properties of the linear and nonlinear GARCH models, both in-sample and out-of-sample volatility forecasts one and ten days ahead were computed. The one-day forecasts can be given in closed form for both models, and for the ten day period the forecasts from the STGARCH models were obtained by Monte Carlo simulation. The conditional variance computed from the estimated models was used as the initial value. The root mean squared errors (RMSE) of forecasts are presented in Table 2. Evaluating volatility forecasts is difficult because true volatility

is not observable as was recently forcefully pointed out by Andersen and Bollerslev (1998), who suggest estimating true conditional volatility from intradaily returns. Because we are here mainly interested only in comparing the performance of the two models, we have chosen to follow the common practice in the literature of approximating the true volatility by squared daily returns. This should be adequate to bring out potential differences between the linear and nonlinear GARCH models. The out-of-sample forecasts were computed for the latter half of the exchange rate return series from STGARCH(1,1) models estimated from the first half. The STGARCH models produce smaller RMSEs in each experiment. For the DEM returns the differences are rather small (less than 0.5%) but for the JPY returns they are more substantial in favor of the STGARCH models, especially in the out-of-sample experiments, where the resulting reductions in RMSEs were 6.7 and 7.5 per cents for the one and ten-day horizons, respectively. This was also to be expected based on the fact that, according to the NICs and IRFs, the properties of the two models are particularly different when the conditional volatility is high, which is the case for the JPY return series in the latter half of the sample period. There the STGARCH model manages to moderate the volatility consequences of shocks while the GARCH model presumably exaggerates persistence. For the DEM return series, on the other hand, the latter half of the sample period is relatively tranquil, and the differences in overall forecast accuracy therefore smaller.

Further evidence for the conjecture that the differences in predictive accuracy mainly arise in periods when the conditional variance is initially high, is provided by computing the conditional RMSEs for different levels of the initial conditional variance implied by the models. In each case the difference in predictive accuracy between the STGARCH model and the corresponding GARCH model turned out to be the more in favor of the former the higher the initial conditional variance was. For the JPY returns the differences in the RMSE in the out-of-sample experiment at the ten-day horizon were 10.6, -4.3 and -35.7 per cents for "low" (lower than 0.3), "middle" (0.3–0.9) and "high" (higher than 0.9) initial variance, respectively, while

the corresponding figures for the DEM returns were 0.1, 2.0 and -41.7 per cents, respectively. This shows that even for the DEM return series for which the overall forecasting performance of the STGARCH model did not seem to much exceed that of the GARCH model, substantial gains could be made when forecasting in periods of high volatility.

6 Conclusion

In this paper we have studied new nonlinear GARCH models motivated by the desire to model time series with highly persistent volatility. Time series of this kind have often proved difficult for conventional GARCH models. The main emphasis was on models that are similar to previous STGARCH models except for the novel feature that a lagged value of conditional variance is used as the transition variable. Empirical applications to exchange rate return series showed that the new STGARCH models can be superior to conventional GARCH models especially in longer term forecasting.

In the paper we also gave sufficient conditions for the stationarity and existence of moments of STGARCH models and some more general nonlinear GARCH models. These results appear quite satisfactory in the case of first order STGARCH models. However, for higher order STGARCH models the situation turned out to be more difficult. Extending these results to higher order models, as well as to presently excluded models with more than a single ARCH term, is therefore an interesting future research problem. Similar extensions to models which allow for dynamics in the conditional expectation are also of interest. A further open research problem related to the considered nonlinear GARCH models is the development of asymptotic theory of estimation and statistical inference. On the applied side, it would be useful to gather more experience about the practical usefulness of the new STGARCH models.

Appendix: Proof of Theorem 1

The idea of the proof is to apply Theorems 15.0.1, 16.01 and 16.15 of Meyn and Tweedie (1993). This requires showing first that the Markov chain X_t defined in (8) is irreducible and aperiodic. The identification of appropriate small sets is also needed. After these preliminaries, we have to verify an appropriate version of the so-called drift condition given in inequality (15.3) of the same reference.

Irreducibility. We start with establishing the irreducibility which is here a somewhat more complicated task than in previous models which only involve ARCH-type conditional heteroskedasticity (cf. e.g. Lu (1996, 1998) and Lu and Jiang (2001). The reason is that the state space of our Markov chain is $\mathcal{X} \subset \mathbb{R}_+^p$ instead of $\mathbb{R}^p$.

To establish the irreducibility of X_t, we need to show that, for some measure φ on the Borel sets of $\mathcal{X}$,

$$\sum_{t=1}^{\infty} P^t(x, A) > 0 \quad \text{for all } x \in \mathcal{X} \tag{16}$$

whenever $\varphi(A) > 0$ (see Meyn and Tweedie (1993, p. 87)). Recall that here $P^t(x, A)$ is the t-step transition probability measure of the Markov chain X_t. In the proof of the following lemma we need an explicit expression for $P^p(x, A)$. To this end, recall from Section 2 that the conditional density function of h_t given $X_{t-1} = x = [x_1 \cdots x_p]'$ is $\psi(\cdot - g(x); x_1)$ where the function $\psi(\cdot;\cdot)$ is defined in equation (9). Using these facts it is straightforward to check that

$$P^p(x, A) = \int_A \prod_{i=0}^{p-1} \psi_i(z; x)\, dz \tag{17}$$

where now $z = [z_1 \cdots z_p]'$,

$$\psi_0(z; x) = \psi(z_p - g(x_1, ..., x_p); x_1)$$

and

$$\psi_i(z; x) = \psi(z_{p-i} - g(z_{p-i+1}, ..., z_p, x_1, ..., x_{p-i}); z_{p-i+1}), \quad i = 1, ..., p-1$$

(cf. the proof of Proposition A1.7 in Chan (1990), or the proof of Lemma 1 in Lu (1998)). For later purposes we also note that the function $\psi(\cdot;\cdot)$ is continuous and positive whenever the first argument is positive.

The following lemma shows the irreducibility of X_t. We use the symbol μ to signify the Lebesgue measure on $\mathbb{R}^p$.

Lemma 5 *Suppose that the assumptions of Theorem 1 hold. Then there exist intervals* $(b_i, c_i) \subset \mathbb{R}_+$, $i = 1, ..., p$, *such that* $D = (b_1, c_1) \times \cdots \times (b_p, c_p) \subset \mathcal{X}$ *and the Markov chain* X_t *is* φ*-irreducible with* $\varphi(\cdot) = \mu(\cdot \cap D)$.

Proof. Assumption 2(ii) implies that, for every $\epsilon > 0$, we can find a real number M_ϵ such that

$$g(x) \leq \sum_{i=1}^{p} (a_i + \epsilon) x_i \quad \text{for } x \in \mathcal{X} \text{ and } |x| > M_\epsilon. \tag{18}$$

Let $\kappa = [\kappa_1 \ \cdots \ \kappa_1]'$ be a vector with $0 < \kappa_i \leq 1$ for all $i = 1, ..., p$. An explicit definition of this vector will be given shortly. Here we use it to define the set

$$S_{1\epsilon} = \{x \in \mathcal{X} : |\kappa x| > M_\epsilon\}. \tag{19}$$

Notice that $|\kappa x| < |x|$ and therefore $|x| > M_\epsilon$ for every $x \in S_{1\epsilon}$. In what follows we also make use of the facts that $|x| = x$ and $|\kappa x| = \kappa x$ for $x \in \mathcal{X}$.

Now, suppose that $X_{t-1} \in S_{1\epsilon}$ and, for simplicity, denote $a_{\epsilon i} = a_i + \epsilon$ $(i = 1, ..., p)$. Then equations (8) and (18) in conjunction with Assumption 1(iii) show that

$$\begin{aligned} \kappa' X_t &= \kappa_1 g(X_{t-1}) + \sum_{i=1}^{p-1} \kappa_{i+1} h_{t-i} + \kappa_1 f(h_{t-1}^{1/2} \varepsilon_{t-1}) \\ &\leq \kappa_1 \sum_{i=1}^{p} a_{\epsilon i} h_{t-i} + \sum_{i=1}^{p-1} \kappa_{i+1} h_{t-i} + \kappa_1 h_{t-1} b(\varepsilon_{t-1}). \end{aligned} \tag{20}$$

Define the matrix $C_\epsilon(\varepsilon_{t-1}) = A_\epsilon + B(\varepsilon_{t-1})$ where the matrix A_ϵ $(p \times p)$ is defined by replacing a_i in (12) by $a_{\epsilon i}$ $(i = 1, ..., p)$ and $B(\varepsilon_{t-1})$ is as in (13). Then the preceding inequality can be written as

$$\kappa' X_t \leq \kappa' C_\epsilon(\varepsilon_{t-1}) X_{t-1}.$$

Note that the quantities on both sides of this inequality are positive. Thus, since $|\kappa' X_t|^k = (\kappa' X_t)^{\otimes k}$, well-known properties of Kronecker's product give

$$\begin{aligned} |\kappa' X_t|^k &\leq (\kappa^{\otimes k})' C_\epsilon (\varepsilon_{t-1})^{\otimes k} X_{t-1}^{\otimes k} \\ &= (\kappa^{\otimes k})' E\left[C_\epsilon (\varepsilon_{t-1})^{\otimes k}\right] X_{t-1}^{\otimes k} + (\kappa^{\otimes k})' \left\{C_\epsilon (\varepsilon_{t-1})^{\otimes k} - E\left[C_\epsilon (\varepsilon_{t-1})^{\otimes k}\right]\right\} X_{t-1}^{\otimes k}. \end{aligned} \tag{21}$$

Using (21) we can show that there exists a finite integer n such that

$$P^n(x, S_{2\epsilon}) > 0, \quad x \in S_{1\epsilon}, \tag{22}$$

where $S_{2\epsilon} = \{x \in \mathcal{X} : \overline{\eta}^k \leq |\kappa' x|^k \leq M_\epsilon\}$ and $\overline{\eta} = \eta \sum_{i=1}^p \kappa_i > 0$ with η given in Assumption 2(i). Clearly, $\mathcal{X} = S_{1\epsilon} \cup S_{2\epsilon}$.

To justify (22), assume that $X_0 = x \in S_{1\epsilon}$ and consider the event

$$\Omega_n = \{a_{\epsilon 1} + b(\varepsilon_{t-1}) < E(a_{\epsilon 1} + b(\varepsilon_{t-1})),\ t = 1, ..., n\}.$$

Due to the nonconstancy of the function b, the probability of the event Ω_n is positive for every value of n. From the definitions it can be seen that the random elements of the matrix $C_\epsilon(\varepsilon_{t-1})$ are of the form $c(a_{\epsilon 1} + b(\varepsilon_{t-1}))^j$ where $j \in \{1, ..., k\}$ and $c > 0$. Furthermore, on Ω_n we have $c(a_{\epsilon 1} + b(\varepsilon_{t-1}))^j < cE(a_{\epsilon 1} + b(\varepsilon_{t-1}))^j$, $j \in \{1, ..., k\}$. Thus, since Assumption 2(ii) entails $h_t \geq \eta > 0$ and since the components of the vector κ are positive it follows from (21) that, on the event Ω_n,

$$|\kappa' X_t|^k \leq (\kappa^{\otimes k})' E\left[C_\epsilon (\varepsilon_{t-1})^{\otimes k}\right] X_{t-1}^{\otimes k}. \tag{23}$$

Now, by assumption we have $\rho\left(E(C_\epsilon(\varepsilon_{t-1}))^{\otimes k}\right) < 1$ when $\epsilon = 0$. Using the continuity of the spectral radius and dominated convergence we can conclude from this that there exists a choice of ϵ such that $\rho\left(E(C_\epsilon(\varepsilon_{t-1}))^{\otimes k}\right) < 1$. (The use of dominated convergence is justified by the assumption $b(x) \leq bx^2$ imposed in Assumption 1(iii).) Such a choice of ϵ will be assumed in all subsequent derivations. Thus, we can use Lemma A.2 of Liang and McAleer (in press) and choose the vector κ such that the vector $\nu = \left(I_{pk} - E(C_\epsilon(\varepsilon_{t-1}))^{\otimes k}\right)' \kappa^{\otimes k}$ has positive components. (The given proof makes clear that it means no loss of generality to assume that the components of κ

are bounded by unity although this is not stated explicitly in the formulation of the lemma.) Thus, we can write inequality (23) as

$$\begin{aligned} \left|\kappa' X_t\right|^k &\leq \left|\kappa' X_{t-1}\right|^k - \nu' X_{t-1}^{\otimes k} \\ &= \left|\kappa' X_{t-1}\right|^k \left(1 - \frac{\nu' X_{t-1}^{\otimes k}}{(\kappa^{\otimes k})' X_{t-1}^{\otimes k}}\right) \\ &\leq (1-\overline{\nu}) \left|\kappa' X_{t-1}\right|^k \end{aligned} \tag{24}$$

where $\overline{\nu} > 0$ is the minimum of the components of ν. Note that here we have also made use of the fact that the components of κ are bounded by unity. Thus, starting with $t = 1$ and proceeding inductively we find from this inequality that, whenever $X_{t-1} \in S_{1\epsilon}$ $(t = 1, ..., n)$,

$$\left|\kappa' X_n\right|^k \leq (1-\overline{\nu})^n \left|\kappa' x\right|^k. \tag{25}$$

Because $0 < 1 - \overline{\nu} < 1$, the right hand side of (25) is smaller that M_ϵ for all n large enough. Choosing n as the smallest integer for which this occurs shows that (22) holds.

Next suppose that $x \in S_{2\epsilon}$. Since the function g is bounded on bounded subsets of $\mathcal{X}$ there exists a positive real number M_p such that $g(x) \leq M_p < \infty$ for all $x \in S_{2\epsilon}$. Without loss of generality, we can assume that $M_p > M_\epsilon$. Thus, we can find an open interval (b_p, c_p) such that $M_p < b_p < c_p < \infty$ and

$$z_p - g(x) > 0, \quad z_p \in (b_p, c_p),\ x \in S_{2\epsilon}.$$

In the same way, when $x \in S_{2\epsilon}$ we have $\sup_{z \in (b_p, c_p)} g(z, x_1, ..., x_{p-1}) \leq M_{p-1} < \infty$ for all $x \in S_{2\epsilon}$. Hence, it is possible to find an open interval (b_{p-1}, c_{p-1}) such that $M_{p-1} < b_{p-1} < c_{p-1} < \infty$ and

$$z_{p-1} - g(z_p, x_1, ..., x_{p-1}) > 0, \quad z_p \in (b_p, c_p),\ z_{p-1} \in (b_{p-1}, c_{p-1}),\ x \in S_{2\epsilon}.$$

Continuing in this way we can find the remaining intervals (b_i, c_i) and use them to construct the set $D = (b_1, c_1) \times \cdots \times (b_p, c_p)$. Then, since the function $\psi(\cdot;\cdot)$ is positive when the first argument is positive, the integrand on the right hand side of

(17) is positive on the set D. Thus, we can conclude that $P^p(x, A \cap D) > 0$ whenever $x \in S_{2\epsilon}$ and for every Borel set A such that $\mu(A \cap D) > 0$. This shows that the desired condition (16) holds for $x \in S_{2\epsilon}$. To show that it also holds for $x \in S_{1\epsilon}$ and to facilitate later proofs, we establish the stronger result

$$\inf_{x \in K} P^p(x, A \cap D) > 0 \tag{26}$$

for every compact set $K \subset S_{2\epsilon}$ and all Borel measurable A such that $\mu(A \cap D) > 0$.

A proof of (26) will be given shortly. Assume for the moment that this result holds and let A be a Borel set such that $\mu(A \cap D) > 0$. Then, by the Chapman-Kolmogorov equation (see Meyn and Tweedie (1993, p. 67)) one obtains, for any $x \in S_{1\epsilon}$,

$$\begin{aligned} P^{n+p}(x, A \cap D) &= \int_{\mathcal{X}} P^n(x, dy) P^p(y, A \cap D) \\ &\geq \int_{S_{2\epsilon}} P^n(x, dy) P^p(y, A \cap D) \\ &\geq cP^n(x, S_{2\epsilon}) \end{aligned} \tag{27}$$

where $c > 0$. Here the former inequality is obvious whereas the latter is obtained from (26) with $K = S_{2\epsilon}$. In view of (22) the last probability in (27) is positive which shows that the condition in (16) also holds for $x \in S_{1\epsilon}$ and $\varphi(\cdot) = \mu(\cdot \cap D)$. Thus, since $\mathcal{X} = S_{1\epsilon} \cup S_{2\epsilon}$, condition (16) holds for all $x \in \mathcal{X}$ with $\varphi(\cdot) = \mu(\cdot \cap D)$.

To complete the proof we still have to establish (26). This can be done by following the arguments in the proof of Lemma 2 of Lu (1998). Thus, suppose that (26) does not hold so that there exists a Borel set A such that $\mu(A \cap D) > 0$ and

$$\inf_{x \in K} P^p(x, A \cap D) = 0, \quad K \subset S_{2\epsilon}.$$

By the definition of infimum, there then exists a sequence $\{\overline{x}_n\} \subset K$ such that $\lim_{n\to\infty} P^p(\overline{x}_n, A \cap D) = 0$. Hence, (17) and Fatou's lemma give

$$\int_{A \cap D} \liminf_{n\to\infty} \prod_{i=0}^{p-1} \psi_i(z; \overline{x}_n)\, dz \leq \lim_{n\to\infty} P^p(\overline{x}_n; A \cap D) = 0$$

We shall obtain a contradiction with this by showing that

$$a(z) \overset{def}{=} \liminf_{n\to\infty} \prod_{i=0}^{p-1} \psi_i(z; \overline{x}_n) > 0$$

for any $z \in A \cap D$.

Let $z \in A \cap D$ be arbitrary but fixed. Then, there exists a subsequence $\{\overline{x}_{n_l}\}$ of $\{\overline{x}_n\}$ such that

$$a(z) = \lim_{l \to \infty} \prod_{i=0}^{p-1} \psi_i(z; \overline{x}_{n_l}). \tag{28}$$

Since $\{\overline{x}_{n_l}\} \subset K$ and K is compact there is a subsequence of $\{\overline{x}_{n_l}\}$, still denoted by $\{\overline{x}_{n_l}\}$, such that $\lim_{l \to \infty} \overline{x}_{n_l} = \overline{x}_0$ for some $\overline{x}_0 \in K$. Let $\overline{x}_{in_l}$ and $\overline{x}_{i0}$ $(i = 1, ..., p)$ signify the components of the vectors $\overline{x}_{n_l}$ and $\overline{x}_0$, respectively. Then, since the function g is bounded on bounded subsets of $\mathbb{R}^p_+$ the limit

$$\lim_{l \to \infty} g(z_{p-i+1}, ..., z_p, \overline{x}_{1n_l}, ..., \overline{x}_{p-i,n_l}) = L_i$$

exists and is finite. (Here L_0 is defined by omitting the arguments $z_{p-i+1}, ..., z_p$.) Thus, combining this fact with (28), the definition of $\psi_i(z; \overline{x}_{n_l})$, and the continuity of the function $\psi(\cdot;\cdot)$ we find that

$$a(z) = \psi_0(z_p - L_0; \overline{x}_{10}) \prod_{i=1}^{p-1} \psi_i(z_{p-i} - L_i; z_{p-i+1}).$$

Now recall that we could assume $M_p > M_\epsilon$ so that $D \subset S_{1\epsilon}$. Thus, since $z \in D$ and $\overline{x}_0 \in K \subset S_{2\epsilon}$ we have $z_{p-i} - L_i > 0$ for all $i = 0, ..., p-1$, as seen from the discussion leading to (26). Since the function $\psi(\cdot;\cdot)$ is positive when the first argument is positive the above equality and the definition of the functions $\psi_i(\cdot;\cdot)$ $(i = 0, ..., p-1)$ yield $a(z) > 0$. Notice that here we also used the assumption $\inf_{x \in \mathbb{R}^p_+} g(x) = \eta > 0$ to ensure that $\psi_0(z_p - L_0; \overline{x}_{10}) > 0$ (cf. Remark 1 of Lu (1998)). This gives the desired contradiction and completes the proof. ■

Aperiodicity and small sets. The following result will be used to prove the aperiodicity of the Markov chain X_t and to identify small sets.

Lemma 6 *Suppose that the assumptions of Theorem 1 hold and let K be any compact subset of $\mathcal{X}$. Then, $\inf_{x \in K} \sum_{j=1}^m P^j(x, A) > 0$ for some positive integer $m = m(A)$ and for all Borel sets A such that $A \subset D$ and $\varphi(A) > 0$.*

Proof. If $K \subset S_{2\epsilon}$ the stated conclusion follows directly from (26). Next suppose that $K \subset S_{1\epsilon}$ and note that we can strengthen (22) to the form $\inf_{x\in K} P^n(x, S_{2\epsilon}) > 0$. To justify this, use the proof given for (22) and dominate the right hand side of (25) by $(1-\overline{\nu}_\epsilon)^n \sup_{x\in K} \kappa' x \leq c < \infty$. Finally, suppose that both $K \cap S_{1\epsilon}$ and $K \cap S_{2\epsilon}$ are nonempty and let n be as in (22). Then, since $\mathcal{X} = S_{1\epsilon} \cup S_{2\epsilon}$, the stated result follows from

$$\inf_{x\in K} \sum_{j=1}^{n+p} P^j(x, A) \geq \min\left\{ \inf_{x\in K\cap S_{1\epsilon}} P^p(x, S_{1\epsilon}), \inf_{x\in K\cap S_{2\epsilon}} P^n(x, S_{2\epsilon}) \right\}.$$

■

The following lemma shows the aperiodicity of the Markov chain X_t.

Lemma 7 *If the assumptions of Theorem 1 hold then the Markov chain X_t is aperiodic.*

Proof. Let the integer $d \geq 1$ stand for the period of X_t and φ_M a maximal irreducibility measure of X_t (for a definition, see Meyn and Tweedie (1993, p. 88-89)). Then, we can partition the state space of X_t as $\mathcal{X} = F \cup E_1 \cup \cdots \cup E_d$ where F, $E_1, ..., E_d$ are disjoint Borel sets such that $\varphi_M(F) = 0$ and $P(x, E_{i+1}) = 1$ for all $x \in E_i$, $i = 0, ..., d-1 \pmod d$ (see Meyn and Tweedie (1993, p. 117)). Assume that X_t is periodic so that $d > 1$. Let D and φ be as in Lemma 5 and notice that, since φ_M is a maximal irreducibility measure, $\varphi_M(F) = 0$ implies $\varphi(F) = 0$. Since $\varphi(D) > 0$ we therefore have $\varphi(E_j \cap D) > 0$ for some $j \in \{1, ..., d\}$. Let $z_j \in E_j$ and $z_{j+1} \in E_{j+1}$. From the assumed periodicity it then follows that $P^n(z_j, E_j \cap D) = 0$ for every n that is not a multiple of d and $P^n(z_{j+1}, E_j \cap D) = 0$ for every n that is a multiple of d. Thus, we have found a compact set $K = \{z_j, z_{j+1}\}$ such that

$$\inf_{x\in K} P^n(x, E_j \cap D) = 0$$

for every positive integer n. This, however, contradicts the result of Lemma 6 because $\varphi(E_j \cap D) > 0$. Thus, we must have $d = 1$. ■

Lemma 8 *If the assumptions of Theorem 1 hold then all compact subsets of $\mathcal{X}$ with positive Lebesgue measure are small.*

Proof. The proof is obtained by applying Lemmas 6 and 7 in conjunction with Proposition A1.3 of Chan (1990). ■

Verification of the drift condition. We know now that the Markov chain X_t is irreducible and aperiodic and that all compact subsets of $\mathcal{X}$ with positive Lebesgue measure are small. Thus, to prove the first statement in part (a) of the theorem, we need to show that condition (15.3) of Meyn and Tweedie (1993, p. 355) holds. To this end, we introduce the function

$$q(x) = 1 + |\kappa' x|^k$$

where κ is the vector defined in the proof of Lemma 5.

We have to consider the conditional expectation

$$E(q(X_t) \mid X_{t-1} = x) = 1 + E\left(\kappa_1 g(x) + \sum_{i=1}^{p-1} \kappa_{i+1} x_i + \kappa_1 f(x_1 \varepsilon_{t-1})\right)^k$$

where the equality is obtained from (8) in the same way as its analog in (20). Now suppose that $x \in S_{1\epsilon}$ defined in (19). Then, the above equality and arguments used in (20) yield

$$\begin{aligned} E(q(X_t) \mid X_{t-1} = x) &\leq 1 + E\left(\kappa_1 \sum_{i=1}^{p} a_{\epsilon i} h_i + \sum_{i=1}^{p-1} \kappa_{i+1} x_i + \kappa_1 x_1 b(\varepsilon_{t-1})\right)^k \\ &= 1 + (\kappa^{\otimes k})' E\left[C_\epsilon(\varepsilon_{t-1})^{\otimes k}\right] x^{\otimes k} \end{aligned}$$

where the equality is obtained by arguments similar to those used to obtain (21) from (20). Repeating the arguments used to arrive at (24) we can now further conclude that

$$\begin{aligned} E(q(X_t) \mid X_{t-1} = x) &\leq q(x)\left[1 - \frac{\nu' x^{\otimes k}}{1 + (\kappa^{\otimes k})' x^{\otimes k}}\right] \\ &\leq (1 - \overline{\nu})\, q(x), \quad x \in S_{1\epsilon}, \end{aligned} \tag{29}$$

where $0 < \overline{\nu} < 1$. On the other hand, by the definition of the set $S_{2\epsilon}$ (see (22)) it is obvious that

$$E(q(X_t) \mid X_{t-1} = x) \leq 1 + |\kappa' M_\epsilon|^k, \quad x \in S_{2\epsilon}. \tag{30}$$

Since $S_{1\epsilon} \cup S_{2\epsilon} = \mathcal{X}$ and the set $S_{2\epsilon}$ is compact it follows from (29) and (30) that condition (15.3) of Meyn and Tweedie (1993, p. 355) holds with the function $q(x)$. Thus, the Markov chain X_t is V-geometrically ergodic with $V(x) = q(x)$. Since $q(x) \geq \overline{\kappa} |x|^k$ with $\overline{\kappa} > 0$ the minimum of the components of the vector κ the stated $|x|^k$-geometric ergodicity of X_t also follows.

The second statement can be proved in the same way as Proposition 4 of Carrasco and Chen (in press) which shows a similar result with $|x|^k$-geometric ergodicity replaced by geometric ergodicity. The only change we have to make in the proof of that proposition concerns statements similar to those in (11). In these statements we have to replace the upper bound unity in supremums by $|x|^k$. By the established $|x|^k$-geometric ergodicity this replacement is justified.

To prove part (b), first recall that we have shown that the Markov chain X_t is irreducible and aperiodic and that condition (15.3) of Meyn and Tweedie (1993, p. 355) holds with an appropriate small set and the function V given by $V(x) = q(x) = 1 + |\kappa' x|^k$. Since it is clear that $c^{-1} |x|^k \leq q(x) \leq c |x|^k$ for some $c \geq 1$ we can therefore conclude from Theorem 16.0.1 of Meyn and Tweedie (1993) that the Markov chain X_t is $|x|^k$-uniformly ergodic (for a definition of this concept, see page 382 of the same reference). Given the $|x|^k$-uniform ergodicity, the stated strong mixing property of X_t follows from Theorem 16.1.5 Meyn and Tweedie (1993, p. 388) and the discussion given after its proof.

Regarding part (c) of the theorem, it suffices to note that geometric ergodicity is known to imply β-mixing in the case of stationary initial values (see Pham (1986) or Doukham (1994, p. 4 and 89)). This completes the proof of Theorem 1.■

References

Andersen, T.G. and T. Bollerslev (1998), Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. International Economic Review 39, 885-905.

Baillie, R.T., T. Bollerslev and H.O. Mikkelsen (1996), Fractionally integrated generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 74, 3-30.

Beine, M. and S. Laurent (2001), Structural change and long memory in volatility: New evidence from daily exchange rates. In "Developments in forecast combination and portfolio choice" by Dunis, C., A. Timmermann and J. Moody (eds.). Wiley, New York.

Bollerslev, T. (1986), Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31, 307-327.

Bougerol, P. and N. Picard (1992), Stationarity of GARCH processes and some non-negative time series. Journal of Econometrics 52, 115-127.

Cai, J. (1994), A Markov model of switching-regime ARCH. Journal of Business & Economic Statistics 12, 309-316.

Carrasco, M. and X. Chen (in press), Mixing and moment properties of various GARCH and stochastic volatility models. Econometric Theory.

Chan, K.S. (1990), Deterministic stability, stochastic stability, and ergodicity. Appendix 1 in H. Tong (1990) "Non-linear Time Series: A Dynamical System Approach". Claredon Press, Oxford.

Doukham, P. (1994), Mixing. Properties and Examples. Springer-Verlag, New York.

Engle, R.F. and C. Mustafa (1992), Implied ARCH models from options prices. Journal of Econometrics 52, 289-311.

Franses, P.H. and D. van Dijk (2000), Non-linear Time Series Models in Empirical Finance. Cambridge U.P., Cambridge, U.K.

Gallant, A.R., P.E. Rossi and G. Tauchen (1993), Nonlinear dynamic structures. Econometrica 61, 871-908.

Glosten, L.R., R. Jagannathan and D. Runkle (1993), Relationship between the expected value and the volatility of the nominal excess return on stocks. Journal of Finance 48, 1779-1801.

González-Rivera, G. (1998), Smooth transition GARCH models. Studies in Nonlinear Dynamics and Econometrics 3, 61-78.

Gray, S.F. (1996), Modeling the conditional distribution of interest rates as a regime-switching process. Journal of Financial Economics 42, 27-62.

Hagerud, G.E. (1997), A New Non-linear GARCH Model. EFI Economics Research Institute, Stockholm.

Hamilton, J.D. and R. Susmel (1994), Autoregressive conditional heteroskedasticity and changes in regime. Journal of Econometrics 64, 307-333.

Klaassen, F. (in press), Improving GARCH Volatility Forecasts with Regime-Switching GARCH. Empirical Economics.

Lamoureux, C.G. and W.D. Lastrapes (1990), Persistence in variance, structural change and the GARCH model. Journal of Business and Economic Statistics 8, 225-234.

Lamoureux, C.G. and W.D. Lastrapes (1993), Forecasting stock return variance: Toward an understanding of stochastic implied volatilities. Review of Financial Studies 5, 293-326.

Lanne, M. and P. Saikkonen (2000), Modeling the U.S. short-term interest rate by mixture autoregressive processes. SFB 373 Discussion Paper 76/2000, Humboldt-Universität zu Berlin.

Ling, S. and M. McAleer (in press), Asymptotics theory for a vector ARMA-GARCH model. Econometric Theory.

Ling, S. and M. McAleer (2002), Stationarity and the existence of moments of a family of GARCH processes. Journal of Econometrics 106, 109-117.

Lu, Z. (1996), A note on geometric ergodicity of autoregressive conditional heteroscedasticity (ARCH) model. Statistics and Probability Letters 30, 305-311.

Lu, Z. (1998), On the geometric ergodicity of a non-linear autoregressive model with an autoregressive conditional heteroskedastic term. Statistica Sinica 8, 1205-1217.

Lu, Z. and Z. Jiang (2001), L_1 geometric ergodicity of a multivariate nonlinear AR model with an ARCH term. Statistics and Probability Letters 51, 121-130.

Lundbergh, S. and T. Teräsvirta (1998), Modelling economic high-frequency time series with STAR-STGARCH models. Working Papers in Economics and Finance 291, Stockholm School of Economics.

Lundbergh, S. and T. Teräsvirta (in press), Evaluating GARCH models. Journal of Econometrics.

Masry, E. and D. Tjøstheim (1995). Nonparametric estimation and identification of nonlinear ARCH time series. Econometric Theory 11, 258–289.

Meyn, S. P. and R. L. Tweedie (1993), Markov Chains and Stochastic Stability. Springer-Verlag, New York.

Nelson, D. (1990), Stationarity and persistence in the GARCH(1,1) model. Econometric Theory 6, 318-334.

Pagan, A.R. and G.W. Schwert (1990), Alternative modes for conditional stock volatility. Journal of Econometrics 45, 267-290.

Pham, D.T. (1986). The mixing property of bilinear and generalised random coefficient autoregressive models. Stochastic Processes and their Applications 23, 291-300.

Taylor, S.J. (2000), Consequences for option pricing of a long memory in volatility. Manuscript, Department of Accounting and Finance, Lancaster University, U.K.

Tjøstheim, D. (1990), Non-linear time series and Markov chains. Advances in Applied Probability 22, 587-611.

Wooldridge, J.M. (1990), A unified approach to robust, regression-based specification tests. Econometric Theory 6, 17-43.

Figure 1: The daily percentage returns for the German mark and Japanese yen against U.S. dollar from January 1, 1980 through July 1998.

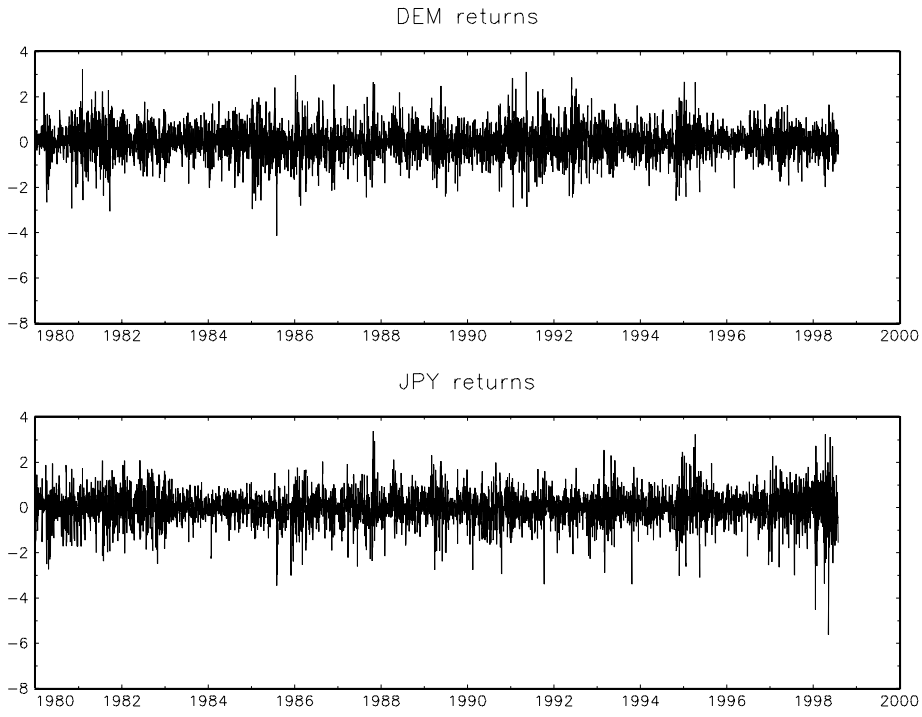

Figure 2: The news impact curves for the STGARCH(1,1) models for the DEM (solid curve) and JPY (dashes) returns as functions of conditional variance h, given $u_t = 1$. The straight lines depict the corresponding news impact curves for the GARCH(1,1) models.

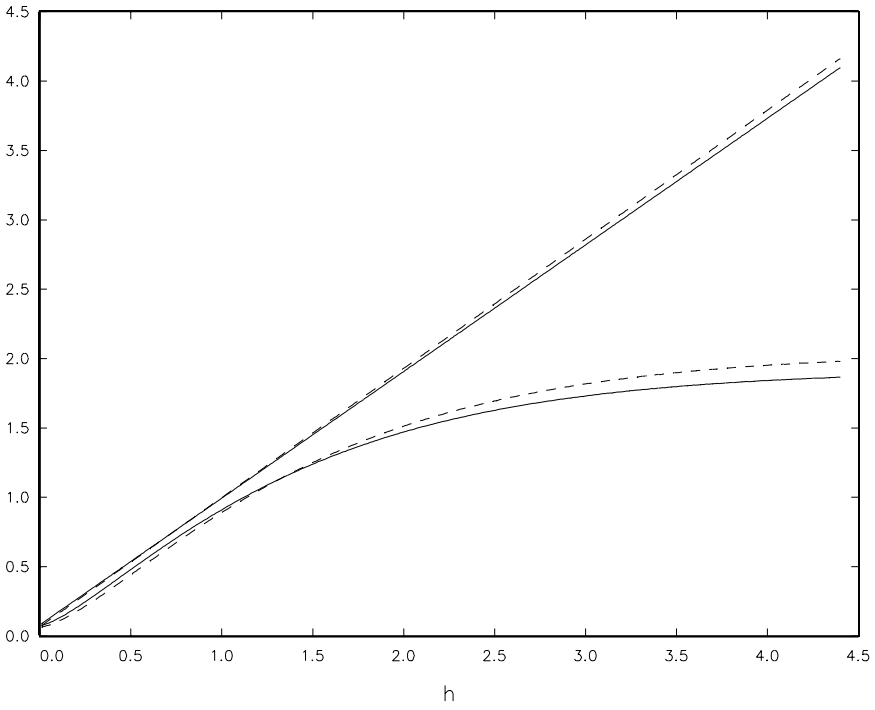

Figure 3: The cumulative impulse response functions for the conditional variance for the estimated STGARCH(1,1) and GARCH(1,1) model for the DEM returns. For STGARCH models (top: $\lambda = 1$, middle: $\lambda = 2$) three levels of initial conditional variance are considered: low ($h_0 = 0.1$, solid line), average ($h_0 = 0.5$, dashes) and high ($h_0 = 2.0$, long dashes). The bottom panel graphs the cumulative impulse response function for the GARCH(1,1) model for $\lambda = 1$.

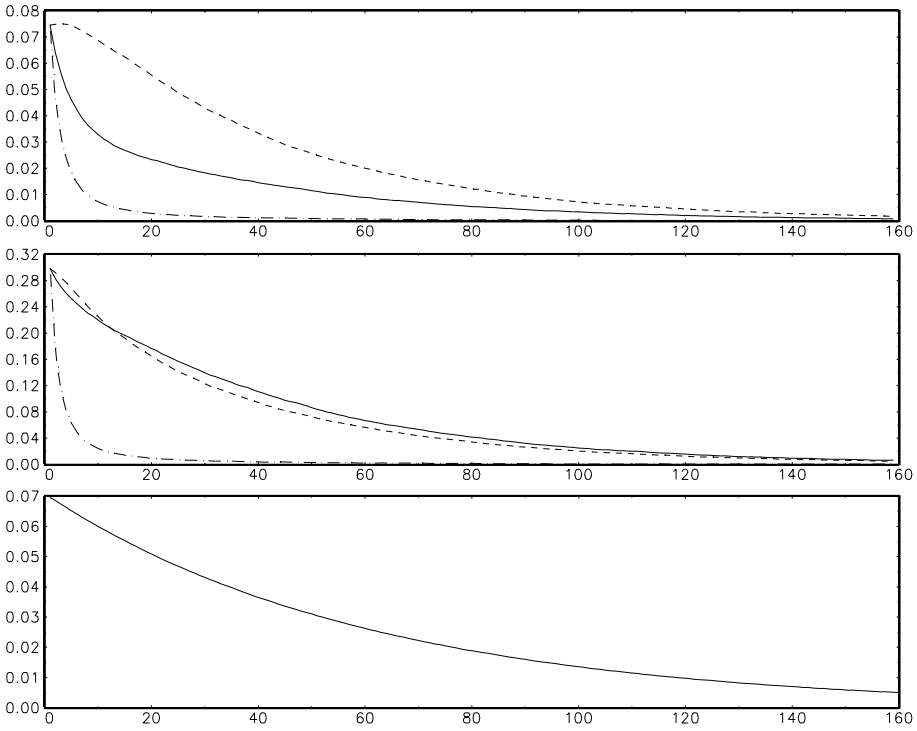

Figure 4: The cumulative impulse response functions for the conditional variance for the estimated STGARCH(1,1) and GARCH(1,1) models for the JPY returns. For STGARCH models (top: $\lambda = 1$, middle: $\lambda = 2$) three levels of initial conditional variance are considered: low ($h_0 = 0.1$, solid line), average ($h_0 = 0.5$, dashes) and high ($h_0 = 2.0$, long dashes). The bottom panel graphs the cumulative impulse response function for the GARCH(1,1) model for $\lambda = 1$.

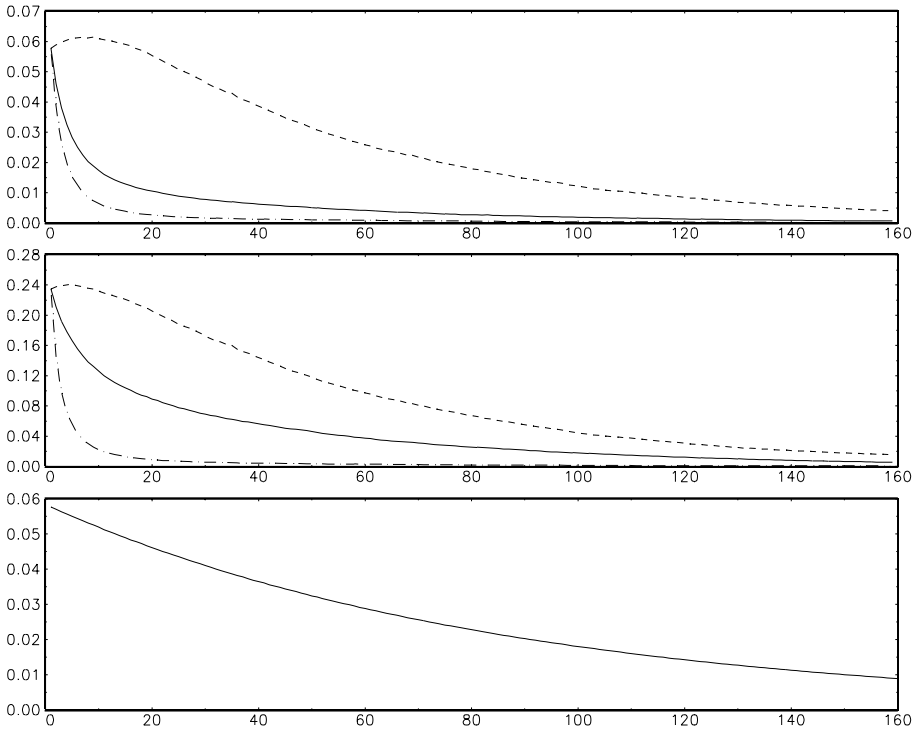

Table 1: Estimation results of the GARCH(1,1)-t and STGARCH(1,1)-t models for the DEM and JPY exchange rate return series.

	DEM		JPY	
	GARCH(1,1)	STGARCH(1,1)	GARCH(1,1)	STGARCH(1,1)
ω	0.009	0.064	0.008	0.083
	(0.002)	(0.012)	(0.002)	(0.013)
α	0.071	0.075	0.058	0.060
	(0.009)	(0.009)	(0.009)	(0.009)
β	0.913		0.931	
	(0.011)		(0.012)	
δ_1		1.843		2.016
		(0.075)		(0.080)
γ		1.434		1.519
		(0.048)		(0.051)
ν	6.318	6.382	4.419	4.457
	(0.592)	(0.603)	(0.302)	(0.305)
AR(1)[a]	0.997	0.956	0.380	0.985
ARCH(10)[b]	0.733	0.426	0.239	0.342
Asymmetry[c]	0.184	0.173	0.006	0.103
Constancy[d]				
$n = 1$	0.159	0.654	0.640	0.348
$n = 2$	0.078	0.145	0.134	0.282
$n = 3$	0.048	0.056	0.223	0.155

The figures in the parentheses are standard errors computed from the inverse of the final Hessian matrix. The figures reported for the diagnostic tests are marginal significance levels.

[a] The alternative model is the corresponding AR(1)-(ST)GARCH model, and under the null hypothesis of no remaining autocorrelation the coefficient of the AR(1) term equals zero. The test is robustified against misspecified conditional variance following Wooldridge (1990, Example 3.3), for details, see Lanne and Saikkonen (2000).

[b] Under the alternative hypothesis ϵ_t in (1) follows an ARCH(10) process, whose coefficients equal zero under the null hypothesis of no remaining ARCH.

[c] Under the alternative hypothesis the coefficient of the ARCH term u_{t-1}^2 depends on the sign of u_{t-1}.

[d] Under the alternative hypothesis the parameters change as a smooth function of time, see Teräsvirta and Lundbergh (in press) for details. Setting $n = 1$ indicates only one-time change, whereas as larger values of n allow for more complicated parameter nonconstancy.

Table 2: Forecast performance of the GARCH(1,1)-t and STGARCH(1,1)-t models for the DEM and JPY exchange rate return series.

	Forecast Horizon	In Sample		Out of Sample	
		GARCH(1,1)	STGARCH(1,1)	GARCH(1,1)	STGARCH(1,1)
DEM	1	0.895	0.880	0.850	0.848
	10	0.930	0.926	0.866	0.861
JPY	1	0.996	0.962	0.971	0.906
	10	0.987	0.959	0.969	0.896

In the out-of-sample experiment the parameters are estimated from data covering the first half of the sample period and held fixed over the forecast period consisting of the latter half of the sample. The figures are root mean squared errors computed as

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(\widehat{h}_t - u_t^2)^2},$$