Meitz, Mika; Saikkonen, Pentti

Working Paper

Ergodicity, mixing, and existence of moments of a class of Markov models with applications to GARCH and ACD models

SSE/EFI Working Paper Series in Economics and Finance, No. 573

Provided in Cooperation with:
EFI - The Economic Research Institute, Stockholm School of Economics

Suggested Citation: Meitz, Mika; Saikkonen, Pentti (2004) : Ergodicity, mixing, and existence of moments of a class of Markov models with applications to GARCH and ACD models, SSE/EFI Working Paper Series in Economics and Finance, No. 573, Stockholm School of Economics, The Economic Research Institute (EFI), Stockholm

This Version is available at:
https://hdl.handle.net/10419/56102

Ergodicity, mixing, and existence of moments of a class of Markov models with applications to GARCH and ACD models*

Mika Meitz
University of Oxford

Pentti Saikkonen
University of Helsinki

SSE/EFI Working Paper Series in Economics and Finance No. 573

First version October 2004, revised December 2005 and April 2007

Abstract

This paper studies a class of Markov models which consist of two components. Typically, one of the components is observable and the other is unobservable or 'hidden'. Conditions under which geometric ergodicity of the unobservable component is inherited by the joint process formed of the two components are given. This implies existence of initial values such that the joint process is strictly stationary and β–mixing. In addition to this, conditions for the existence of moments are also obtained and extensions to the case of nonstationary initial values are provided. All these results are applied to a general model which includes as special cases various first order generalized autoregressive conditional heteroskedasticity (GARCH) and autoregressive conditional duration (ACD) models with possibly complicated nonlinear structures. The results only require mild moment assumptions and in some cases provide necessary and sufficient conditions for geometric ergodicity.

*The first author's work was financially supported by the Jan Wallander's and Tom Hedelius' Foundation, Grant No. J03–41. The second author acknowledges financial support by the Academy of Finland and the Yrjö Jahnsson Foundation. We thank a Co-Editor and three anonymous referees for helpful comments and suggestions. First version of this paper was completed in October 2004 while the first author was a PhD student at Stockholm School of Economics. Subsequently, material from it has been presented at the New Zealand Econometric Study Group Meeting, Christchurch, March 2005, International Workshop on Financial Econometrics and Statistics, Perth, April 2005, RUESG Workshop on Financial Econometrics, Helsinki, August 2005, Econometric Society World Congress, London, August 2005, International Conference on Finance, Copenhagen, September 2005, European Winter Meeting of the Econometric Society, Istanbul, October 2005, and in seminars at Stockholm School of Economics, Université catholique de Louvain, and Victoria University of Wellington. We thank the participants in the occasions, Michael McAleer in particular, for their comments. Address correspondence to: Mika Meitz, Department of Economics, University of Oxford, Manor Road Building, Manor Road, Oxford, OX1 3UQ, United Kingdom; e-mail: mika.meitz@economics.ox.ac.uk; or to: Pentti Saikkonen, Department of Mathematics and Statistics, University of Helsinki, P. O. Box 68, FIN–00014 University of Helsinki, Finland; e-mail: pentti.saikkonen@helsinki.fi.

1 Introduction

This paper is concerned with probabilistic properties of two common classes of models, namely generalized autoregressive conditional heteroskedasticity (GARCH) models and autoregressive conditional duration (ACD) models. GARCH models were pioneered by Engle (1982) and Bollerslev (1986), and have ever since been widely used to analyze financial time series. The more recent ACD models were introduced by Engle and Russell (1998) to model the time dimension of irregularly spaced ultra-high-frequency data.

Our study of GARCH and ACD models makes use of the theory of Markov chains. Both GARCH and ACD models can be thought of as consisting of two components of which one is observable (say, returns or durations) and the other is unobservable or 'hidden' (say, conditional variance or conditional expected duration). From the viewpoint of Markov chain theory, the unobservable component can be investigated as a Markov chain of its own in isolation from the observable component. However, it is also useful to consider both components jointly as a single Markov chain. For instance, in the development of statistical estimation and testing theory it is pertinent to know when the joint process formed of the two components is, for example, stationary and has finite moments of some order. To make such results readily available, we obtain conditions under which the ergodicity, or more precisely, V–geometric ergodicity of the hidden process (viewed as a Markov chain of its own) is inherited by the joint process (consisting of both the observable and hidden components). An immediate consequence of this is that, with an appropriate choice of initial values, the joint process is strictly stationary and β–mixing (or absolutely regular) with certain moments existing. Building on the recent work of Liebscher (2005) we also obtain conditions which imply β–mixing in the case of nonstationary initial values. Because interest in results of this kind may not be confined to GARCH and ACD models they are first obtained for a very general class of Markov models defined in terms of transition probability measures. This extends the work of Liebscher (2005) who obtained unified sufficient conditions for geometric ergodicity and β–mixing of autoregressive models and provided an insightful discussion about their relationship.

We apply the results obtained for our general class of Markov models to a sub-class which contains many first order GARCH and ACD models as special cases. For simplicity,

we concentrate on the leading case of first order GARCH and ACD models but allow for more complicated nonlinear structures than in earlier literature. Our results apply to the families of GARCH and ACD models introduced by Hentschel (1995) and Fernandes and Grammig (2006), respectively, and thereby to several commonly used GARCH and ACD models. Our results also apply to the integrated GARCH (IGARCH) model and provide a rigorous proof of its short memory nature previously demonstrated by Ding and Granger (1996) using more elementary methods. In addition to these models, the GARCH-in-mean (GARCH–M) model as well as some GARCH and ACD models with rather complicated nonlinear structures are also covered. For some of these models the obtained conditions for ergodicity, mixing, and stationarity appear new.

The approach used in this paper has predecessors. Genon-Catalot, Jeantheau, and Larédo (2000) considered a general class of Markov models referred to as a 'hidden Markov model' and obtained results similar to ours for stochastic volatility models. Carrasco and Chen (2002) attempted to generalize these results by formulating a 'generalized hidden Markov model' which could also be applied to GARCH and ACD models. Unfortunately, however, this generalization appears too general to be useful. We show by a counterexample that the conditions required for the generalized hidden Markov model do not necessarily guarantee the validity of the ergodicity and mixing results given by Carrasco and Chen (2002). We wish to stress, however, that this only concerns their results on the generalized hidden Markov model. All their results on GARCH, stochastic volatility, and ACD models still remain valid as long as their results on the generalized hidden Markov model are corrected. Our paper presents such a correction (Proposition 1 below). An alternative correction has recently been provided by Carrasco and Chen (2005).

As far as GARCH and ACD models are concerned, it should be mentioned that related results on ergodicity, mixing, strict stationarity, and existence of moments have also appeared in Nelson (1990), Bougerol and Picard (1992), Duan (1997), Ling (1999), Zhang, Russell, and Tsay (2001), Carrasco and Chen (2002), Ling and McAleer (2002), Ling and McAleer (2003), Lanne and Saikkonen (2005), Lee and Shin (2005), Francq and Zakoïan (2006), Meitz (2006), Straumann and Mikosch (2006), Cline (2006), Fernandes, Medeiros, and Veiga (2006), Medeiros and Veiga (2006), and Kristensen (2007). Our contribution to this work is that we show how these models can be handled in a unified framework

which also applies when very general nonlinear structures or even models, other than GARCH and ACD models, are of interest. Even though our framework is very general it provides necessary and sufficient conditions for geometric ergodicity in some special cases and, similarly to Francq and Zakoïan (2006), only requires mild moment assumptions.

The rest of this paper is organized as follows. Our general class of Markov models is studied in Section 2. In Section 3 these results are specialized to a specific sub-class of models which contains various GARCH and ACD models. Concluding remarks are presented in Section 4. Proofs of all the results are given in an Appendix.

2 General Markov model

2.1 Definition

We motivate our general Markov model by using the standard GARCH(1,1) model

$$\begin{aligned} u_t &= h_t^{1/2}\varepsilon_t \\ h_t &= \omega + \beta h_{t-1} + \alpha u_{t-1}^2, \end{aligned} \tag{1}$$

where $\varepsilon_t \sim$ i.i.d.$(0,1)$ with ε_t independent of (u_s, h_s), $s < t$, and the parameters satisfy $\omega > 0$, $\alpha \geq 0$, $\beta \geq 0$, and $\alpha + \beta < 1$. Here u_t is an observed process and h_t is its unobserved conditional variance. Substituting h_t from the latter equation to the former shows that (u_t, h_t) is a Markov chain. On the other hand, substituting u_t from the former equation to the latter shows that h_t can be viewed as a separate Markov chain defined by the equation $h_t = \omega + \beta h_{t-1} + \alpha h_{t-1}\varepsilon_{t-1}^2$. Thus, one can use the theory of Markov chains to study properties of either the joint process (u_t, h_t) or of h_t in isolation from the process u_t. An approach like this was recently used by Carrasco and Chen (2002) who showed how to extend results on stationarity and mixing obtained for h_t to the joint process (u_t, h_t). These authors also defined a 'generalized hidden Markov model' in order to make their approach generally applicable. We adopt the same idea and consider a general Markov model with a structure similar to that in the GARCH(1,1) model (1). A comprehensive reference of the needed Markov chain theory is Meyn and Tweedie (1993) whereas Chan (1990) provides a short review. As a further reference we mention Doukhan (1994) where the employed concept of β–mixing and its relation to other mixing concepts are discussed.

Consider two stochastic processes, Y_t and X_t $(t = 0, 1, \ldots)$, taking values in measurable spaces $(\mathcal{Y}, \mathcal{B}(\mathcal{Y}))$ and $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$, respectively. Suppose the joint process $Z_t = (Y_t, X_t)$ is a (time homogeneous) Markov chain on $(\mathcal{Z}, \mathcal{B}(\mathcal{Z}))$, where $\mathcal{Z} = \mathcal{Y} \times \mathcal{X}$ and $\mathcal{B}(\mathcal{Z}) = \mathcal{B}(\mathcal{Y} \times \mathcal{X})$, and let $P_Z^n(z, A) = \Pr(Z_n \in A \mid Z_0 = z)$, $z \in \mathcal{Z}$, $A \in \mathcal{B}(\mathcal{Z})$, signify its n–step transition probability measure ($P_Z^1(\cdot, \cdot) = P_Z(\cdot, \cdot)$ and similarly for other transition probability measures). As in the generalized hidden Markov model of Carrasco and Chen (2002) (conditions (ii) and (iii) of their Definition 3) we now assume that, for all $t \geq 1$, the conditional distribution of Y_t given $(X_t, Y_{t-1}, X_{t-1}, \ldots, Y_0, X_0)$ only depends on X_t and that the conditional distribution of Y_t given $X_t = x$ does not depend on t. Let $\pi_{Y|X}(\cdot \mid x)$ signify this conditional probability distribution. As in the proof of Proposition 4(i) of Carrasco and Chen (2002) we can then write $P_Z^n(z, dz) = \Pr(dy \mid dx, Z_0 = z) \Pr(dx \mid Z_0 = z)$ where $z = (y, x)$ and the former factor of the product can be replaced by $\pi_{Y|X}(dy \mid x)$. In the aforementioned proof, Carrasco and Chen (2002) use the assumption that X_t is an (unobserved) stationary Markov chain (condition (i) of their Definition 3) and replace the latter factor by $P_X^n(x, dx)$, the n–step transition probability measure of X_t. However, this replacement is problematic in GARCH models, for example. Although h_t in the GARCH(1,1) model (1) can be treated as a Markov chain of its own this Markov chain is not identical to the latter component of the joint process (u_t, h_t). Specifically, given an initial value (u_0, h_0), the joint process implies that $h_1 = \omega + \beta h_0 + \alpha u_0^2$ whereas $h_1 = \omega + \beta h_0 + \alpha h_0 \varepsilon_0^2$ results when h_t is treated as a separate Markov chain. Thus, if the joint process (u_t, h_t) is the Markov chain considered, the conditional probability distribution of h_1 also depends on the initial value u_0, not only on h_0.

Motivated by the preceding discussion we denote $\widetilde{P}_X^n(z, \cdot) = \Pr(X_n \in \cdot \mid Z_0 = z)$ and conclude that

$$P_Z^n(z, dz) = \pi_{Y|X}(dy \mid x) \, \widetilde{P}_X^n(z, dx). \tag{2}$$

As noticed above, the dependence of the latter factor on the right hand side on the initial value y is inconvenient. Fortunately, however, this matter can be handled (at least) in GARCH models. To see this, consider again the GARCH(1,1) model (1) and the related two-dimensional Markov chain. Given the initial value (u_0, h_0), the joint process implies that $h_1 = \omega + \beta h_0 + \alpha u_0^2$ and, as can be easily checked, $h_2 = \omega + \beta \widetilde{h}_0 + \alpha \widetilde{h}_0 \varepsilon_1^2$ where $\widetilde{h}_0 = \omega + \beta h_0 + \alpha u_0^2$. On the other hand, when h_t is treated as a separate Markov chain,

$h_1 = \omega + \beta h_0 + \alpha h_0 \varepsilon_0^2$ is obtained. Thus, the generation mechanism of h_2 (based on the two-dimensional Markov chain) is entirely similar to that of h_1 obtained when h_t is treated as a separate Markov chain. Only the initial value $\widetilde{h}_0$ that appears in h_2 is defined in a special way. This clearly extends to larger values of t so that, apart from the definition of the initial value, the generation mechanism of h_t $(t \geq 2)$ based on the two-dimensional Markov chain (u_t, h_t) is identical to that of h_{t-1} obtained when h_t is analyzed separately.

Using the above discussion on the GARCH(1,1) model (1) as a pattern we now replace the probability measure $\widetilde{P}_X^n(z, \cdot)$ in (2) by a counterpart which, for some $j \geq 0$, can be treated as an $(n-j)$–step transition probability measure of a separate Markov chain on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$. We state the following assumption.

Assumption 1 *Let $Z_t = (Y_t, X_t)$ $(t = 0, 1, \ldots)$ be a Markov chain on $(\mathcal{Z}, \mathcal{B}(\mathcal{Z}))$ where $\mathcal{Z} = \mathcal{Y} \times \mathcal{X}$ and $\mathcal{B}(\mathcal{Z}) = \mathcal{B}(\mathcal{Y} \times \mathcal{X})$. Assume the following conditions.*

(a) For all $n \geq 1$ the n–step transition probability measure of Z_t can be expressed in the form (2) where $\pi_{Y|X}(\cdot \mid x)$ is the conditional probability distribution of Y_t given $X_t = x$.

(b) There exist a function $\lambda : \mathcal{Z} \to \mathcal{X}$, an integer $j \geq 0$, and a transition probability measure $P_X(\cdot, \cdot)$ of a Markov chain on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ such that, for all $n > j$, $\widetilde{P}_X^n(z, \cdot) = P_X^{n-j}(\widetilde{x}, \cdot)$ where $\widetilde{x} = \widetilde{x}(z) = \lambda(z)$.

It is implicit in Assumption 1(a) that conditions (ii) and (iii) defining the generalized hidden Markov model of Carrasco and Chen (2002) are satisfied. Furthermore, when X_t is viewed as a part of the joint process (Y_t, X_t) its transition probability measure is assumed to agree with the transition probability measure of a separate Markov chain on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ with suitably defined initial values. However, X_t is not necessarily a Markov chain because its transition probabilities may depend on the initial value of the joint process (Y_t, X_t). Therefore, condition (i) of Definition 3 of Carrasco and Chen (2002) need not hold (not even without the word 'stationary').

Assumption 1 is more restrictive than required for the generalized hidden Markov model of Carrasco and Chen (2002). To demonstrate the need of such a restriction, we

use a model which is a special case of Example 1 of Carrasco and Chen (2002). Thus, let ε_t be a sequence of n.i.d.$(0,1)$ random variables and consider the model

$$\begin{aligned} Y_t &= \varepsilon_t + \varepsilon_{2t} \\ X_t &= \varepsilon_t, \end{aligned} \tag{3}$$

$t = 1, 2, \ldots$. The model is extended for $t = 0$ by assuming that Y_0 and X_0 are independent of each other and of $\{\varepsilon_t,\ t \geq 1\}$ with standard normal distributions. Clearly, X_t is a stationary, geometrically ergodic, and β–mixing Markov chain. It is also straightforward to verify that the conditions required for the generalized hidden Markov model of Carrasco and Chen (2002) are satisfied and, by Proposition 4 of that paper, the joint process (Y_t, X_t) should be a geometrically ergodic and β–mixing Markov chain. This, however, is not the case. Because $\mathrm{Cov}(Y_t, Y_{2t}) = \mathrm{Cov}(\varepsilon_{2t}, \varepsilon_{2t}) = 1$ for all $t \geq 1$ the process Y_t is not strong mixing and, hence, not β–mixing (cf. Proposition 1, p. 4, and Theorem 3, p. 9, in Doukhan (1994)). Also, the conditional distribution of (Y_t, X_t) given its past is not a function of (Y_{t-1}, X_{t-1}) only and, therefore, (Y_t, X_t) is not a (geometrically ergodic) Markov chain.

An important requirement in Assumption 1 is that the joint process (Y_t, X_t) is a Markov chain because then models like (3) are ruled out.[1] It is straightforward to check that Assumption 1 holds for the GARCH(1,1) model (1) with $j = 1$, the function λ given by $\lambda(u, h) = \omega + \beta h + \alpha u^2$, $\pi_{Y|X}(\cdot \mid x)$ the conditional distribution of u_t given h_t, and $P_X(\cdot, \cdot)$ the transition probability measure associated with h_t viewed as a separate Markov chain. The GARCH(1,1) model (1) is a special case of the model

$$Y_t = F_y(X_t, \zeta_t) \tag{4}$$

$$X_t = F_x(X_{t-1}, Y_{t-1}), \tag{5}$$

where ζ_t is an i.i.d. error term independent of (Y_s, X_s), $s < t$, and the random vectors Y_t, X_t, and ζ_t take values on some subsets of Euclidean spaces equipped with Borel sigma fields. As in the GARCH(1,1) model (1) it can be seen that the joint process (Y_t, X_t) is a

[1] Recently, Carrasco and Chen (2005) have corrected their original definition of the generalized hidden Markov model in a way slightly different from our Assumption 1. When this corrected definition is used undesirable models such as (3) are ruled out and all the results in Carrasco and Chen (2002), including Proposition 4, remain valid.

Markov chain and that ζ_t and X_t are independent with the latter having the representation

$$X_t = F_x\left(X_{t-1}, F_y(X_{t-1}, \zeta_{t-1})\right) \stackrel{def}{=} G_x(X_{t-1}, \zeta_{t-1}). \tag{6}$$

Thus, X_t can be viewed as a Markov chain of its own and studied in isolation from Y_t. Verifying that Assumption 1 is satisfied for the model (4)–(5) with $\lambda(\cdot) = F_x(\cdot)$ is straightforward but somewhat technical. Details are therefore deferred to the Appendix.

2.2 Geometric ergodicity

We shall now show that Assumption 1 guarantees that ergodicity and mixing of the component process X_t are inherited by the joint process Z_t. We use the V–geometric ergodicity of a Markov chain defined as follows (see Meyn and Tweedie (1993, p. 356)).

Definition 1 *The Markov chain Z_t is V–geometrically ergodic if there exist a real valued function $V : \mathcal{Z} \to [1, \infty)$, a probability measure π_Z on $\mathcal{B}(\mathcal{Z})$, and constants $\varrho < 1$ and $M_z < \infty$ (depending on z) such that*

$$\sup_{v:|v|\leq V} \left| \int_{\mathcal{Z}} P_Z^n(z, dw)\, v(w) - \int_{\mathcal{Z}} \pi_Z(dw) v(w) \right| \leq \varrho^n M_z \quad \textit{for all } z \in \mathcal{Z} \textit{ and all } n \geq 1. \tag{7}$$

The definition also assumes that the function V is integrable with respect to the probability measure π_Z. When condition (7) holds we also say that the transition probability measure $P_Z(\cdot, \cdot)$ is V–geometrically ergodic and similarly for other transition probability measures such as $P_X(\cdot, \cdot)$. Note that the first integral in (7) equals the conditional expectation $E[v(Z_n) \mid Z_0 = z]$.

The weakest form of V–geometric ergodicity is obtained when $V(\cdot) \equiv 1$ in which case the Markov chain Z_t is said to be geometrically ergodic. Geometric ergodicity entails that the n–step transition probability measure $P_Z^n(z, \cdot)$ converges at a geometric rate to the probability measure $\pi_Z(\cdot)$ with respect to the total variation norm for all $z \in \mathcal{Z}$. The probability measure π_Z is often referred to as the stationary probability measure of Z_t. The reason is that geometric ergodicity implies stationarity of Z_t if the initial value Z_0 is distributed according to the probability measure π_Z (see Meyn and Tweedie (1993, p. 230–231)). A convenient feature of V–geometric ergodicity is that it automatically shows the existence of the expectation of $\int_{\mathcal{Z}} \pi_Z(dw) v(w)$ for all v such that $|v(\cdot)| \leq V(\cdot)$.

The following proposition establishes the V–geometric ergodicity of Z_t.

Proposition 1 *Suppose that the Markov chain $Z_t = (Y_t, X_t)$ satisfies Assumption 1 and that the transition probability measure $P_X(\cdot,\cdot)$ is V_X–geometrically ergodic. Then Z_t is V_Z–geometrically ergodic for any function $V_Z : \mathcal{Z} \to [1,\infty)$ such that $\int_{\mathcal{Y}} \pi_{Y|X}(dy \mid x)\, V_Z(y,x) \leq cV_X(x)$ for all $x \in \mathcal{X}$ and some $c < \infty$.*

The condition imposed on the function V_Z in Proposition 1 is automatically satisfied for $V_Z(y,x) = V_X(x)$, although more useful results can be obtained with other choices of V_Z. However, even this special case shows that the geometric ergodicity of $P_X(\cdot,\cdot)$ is inherited by Z_t and, when initialized from its stationary distribution, Z_t is stationary and β–mixing with geometrically decaying mixing numbers (see Meyn and Tweedie (1993, pp. 230–231) and Doukhan (1994, p. 4 and 89)). Thus, Proposition 1 provides us with results similar to those stated in Proposition 4 of Carrasco and Chen (2002).

While Proposition 1 makes it possible to apply limit theorems developed for Markov chains to functions of Z_t regardless of initial values (see Meyn and Tweedie (1993, Chapter 17)), these theorems are not directly applicable when functions of $(Z_t, Z_{t-1}, ..., Z_1)$ are of interest (such functions are met, for instance, in the estimation theory of GARCH models, see e.g. Straumann and Mikosch (2006, Sections 6 and 7)). In such cases other limit theorems can be useful. Our next result gives sufficient conditions for Z_t to be β–mixing. This result holds for a wide class of nonstationary initial values and provides the needed prerequisite for using limit theorems developed for near epoch dependent functions of mixing processes (see Davidson (1994, Sections 20.6 and 24.4) for such limit theorems and Francq and Zakoïan (2006) for a related discussion on the usefulness of limit theorems developed for mixing processes). We use a subscript in the expectation operator to indicate the initial distribution of the chain with respect to which the expectation is taken.

Proposition 2 *Let the assumptions of Proposition 1 be satisfied and the function V_Z be as required in Proposition 1. Furthermore, let $\pi_X(\cdot)$ signify the stationary probability measure related to a Markov chain with transition probability measure $P_X(\cdot,\cdot)$. Suppose*

that (a) there exist constants $\varrho < 1$ and $R < \infty$ such that

$$\sup_{v:|v|\leq V_X} \left| \int_{\mathcal{X}} P_X^n(x, dw)\, v(w) - \int_{\mathcal{X}} \pi_X(dw) v(w) \right| \leq \varrho^n R V_X(x) \tag{8}$$

for all $x \in \mathcal{X}$ and all $n \geq 1$, (b) $E_\mu[V_X(\lambda(X_0, Y_0))] < \infty$ where μ is the distribution of the initial value $Z_0 = (Y_0, X_0)$, and (c) $\int_{\mathcal{Y}} \pi_{Y|X}(dy \mid x)\, V_X(\lambda(x, y)) \leq c V_X(x)$ for all $x \in \mathcal{X}$ and some $c < \infty$. Then Z_t is β–mixing with geometrically decaying mixing numbers.

This proposition is based on recent results of Liebscher (2005) which highlight relations between β–mixing and geometric ergodicity. To be able to apply these results we need assumptions not needed in the case of stationary initial values. Our assumption (a) is slightly stronger than V_X-geometric ergodicity of $P_X(\cdot, \cdot)$, but is implied by the so-called drift criterion which is a standard tool used to obtain geometric ergodicity (see Meyn and Tweedie (1993, Theorem 15.0.1)). A counterpart of our third assumption (c) was already needed in Proposition 1. This assumption is not very restrictive either in that it is automatically satisfied by the general model (4)–(5) and, therefore, by models we are mainly interested in. We state this as a lemma.

Lemma 1 *Condition (c) of Proposition 2 is redundant for the model (4)–(5).*

Assumptions (a) and (c) also imply that Z_t is $V-$geometrically ergodic with $V(\cdot) = V_X(\lambda(\cdot))$ and, for this case, the argument given in Meyn and Tweedie (1993, discussion following Theorem 16.1.5) could be employed to establish the strong mixing of Z_t. This argument assumes condition (b) and suggests that it is also necessary in our case.

It may also be noted that the conditions of Proposition 2 ensure finiteness of certain moments. For instance, it is straightforward to establish that $\sup_{t\geq 1} E_\mu[v(Z_t)] < \infty$ for any function v such that $|v(\cdot)| \leq V_Z(\cdot)$, and that these moments converge to the ones taken with respect to the stationary distribution π_Z at a geometric rate.

3 GARCH and ACD models

As in Carrasco and Chen (2002), Propositions 1 and 2 can be applied to first order GARCH models. In addition to various GARCH(1,1) models Carrasco and Chen (2002)

also considered higher-order GARCH models as well as examples of ACD models and autoregressive stochastic volatility models. We do not consider stochastic volatility models because for them similar results can be found in Genon-Catalot, Jeantheau, and Larédo (2000). However, as an extension of previous work we present both GARCH models and ACD models as special cases of a general model which even includes the GARCH–M model. As far as we know, these are the first results on geometric ergodicity and mixing obtained for the GARCH–M model (for the ARCH–M model similar results were obtained by Masry and Tjøstheim (1995)).

We consider a special case of the model (4)–(5) with Y_t and X_t real valued and X_t positive. Specifically, the model is defined by

$$Y_t = f_{y1}(X_t) + f_{y2}(X_t)\varepsilon_t \tag{9}$$

$$X_t = f_{x1}(X_{t-1}) + f_{x2}(Y_{t-1} - f_{y1}(X_{t-1}), X_{t-1}), \tag{10}$$

where the ε_t are i.i.d. and independent of (Y_s, X_s), $s < t$, and f_{y1}, f_{y2}, f_{x1}, and f_{x2} are Borel measurable functions to be described in detail shortly. The analog of equation (6) is obtained by substituting Y_{t-1} from (9) into (10), yielding

$$X_t = f_{x1}(X_{t-1}) + f_{x2}(f_{y2}(X_{t-1})\varepsilon_{t-1}, X_{t-1}). \tag{11}$$

A model formulated in this way incorporates various GARCH and ACD models. In the GARCH context, f_{y1} is the conditional mean function whereas f_{y2} is used to model the conditional variance. In the ACD context, f_{y2} represents the conditional mean of Y_t and f_{y1} is omitted. Concrete examples will be given later.

For the development of our theory we make the following assumptions.

Assumption 2

(a) The i.i.d. random variables ε_t have a probability density function $\phi_\varepsilon(\cdot)$ supported on $(\underline{\varepsilon}, \infty)$ and bounded away from zero on compact subsets of $(\underline{\varepsilon}, \infty)$. Here either $\underline{\varepsilon} = 0$ or $\underline{\varepsilon} = -\infty$.

(b) The functions $f_{x1} : \mathbb{R}_+ \to \mathbb{R}_+$ and $f_{x2} : (\underline{\varepsilon}, \infty) \times \mathbb{R}_+ \to \mathbb{R}$ are bounded on bounded subsets of their domains and, for some $\underline{f} > 0$, $\inf_{x \in \mathbb{R}_+, u \in (\underline{\varepsilon}, \infty)} (f_{x1}(x) + f_{x2}(u, x)) = \underline{f}$.

(c) There exists a real number $a \in [0, \infty)$ *such that* $f_{x1}(x) \leq ax + o(x)$ *as* $x \to \infty$.

(d) The function f_{x2} *satisfies the following three conditions.*

(d_1*) There exists an unbounded interval of* $\mathbb{R}_+$ *which is, for all* $x > 0$*, contained in the image set* $f_{x2}((\underline{\varepsilon}, \infty), x)$.

(d_2*) For all* $x > 0$*, the function* $f_{x2}(\cdot, x)$ *is continuous from the right (or alternatively, continuous from the left).*

(d_3*) There exists a real number* $R > 0$ *such that, for* $u > R$ *and all* $x > 0$*,* $f_{x2}(u, x)$ *is continuous and monotonically increasing, and the related inverse function* $f_{x2}^{-1}(v, x)$ *has a partial derivative* $\partial f_{x2}^{-1}(v, x)/\partial v$ *which is bounded away from zero on compact subsets of its domain.*

(e) There exists a Borel measurable function $b : (\underline{\varepsilon}, \infty) \to \mathbb{R}_+$*, nonconstant and continuous on some open set, and a real number* $c \in [0, \infty)$ *such that* $f_{x2}(f_{y2}(x)\varepsilon_t, x) \leq xb(\varepsilon_t) + c$ *for all* $x \in \mathbb{R}_+$*. Furthermore,* $E[b(\varepsilon_t)^k] < \infty$ *for some* $k \in \mathbb{R}_+$.

(f) The function $f_{y2} : \mathbb{R}_+ \to \mathbb{R}_+$ *is bounded on bounded subsets of its domain and bounded away from zero on* $[\underline{f}, \infty)$.

Assumption 2(a) is satisfied in most applications. The case $\underline{\varepsilon} = 0$ is typical in ACD models, while in GARCH models $\underline{\varepsilon} = -\infty$. Restricting $\underline{\varepsilon}$ to these two values is not essential for the development of the theory, but from a practical point of view there is little need for generalizing this. In Assumptions 2(b) and (f) the functions f_{x1}, f_{x2}, and f_{y2} are assumed to be bounded on bounded subsets of their domains, while in Assumptions 2(a), (b), (d), and (f) the functions ϕ_ε, $f_{x1} + f_{x2}$, $\partial f_{x2}^{-1}(v, x)/\partial v$, and f_{y2}, respectively, are assumed to be suitably bounded away from zero. These assumptions are hardly restrictive in practice. Especially the latter assumptions, though, suggest that some attention needs to be paid to the definitions of these functions.

Assumption 2(b) also ensures that the process X_t is always positive. Note that there is more than one way to define the functions f_{x1} and f_{x2} without violating equations (10) and (11). In most cases it is natural to choose the functions f_{x1} and f_{x2} such that both

of them are always positive. However, as a subsequent example shows, it is useful to be more flexible and only require that the sum $f_{x1} + f_{x2}$ is positive.

The conditions restricting the functions f_{x1} and f_{x2} in Assumptions 2(c) and (e), respectively, essentially restrict X_t to depend on its past value at most in a linear fashion when arbitrarily large past values are of concern. This will be crucial in proving the geometric ergodicity of X_t. Similar assumptions have also been used in previous proofs for geometric ergodicity (see Lanne and Saikkonen (2005) for both Assumption 2(c) and (e) and Masry and Tjøstheim (1995), Lu (1998), and Lu and Jiang (2001), among others for Assumption 2(c)). Further conditions on the constant a, the function $b(\cdot)$, and moments of the random variables ε_t will be imposed later. It will prove beneficial to have the values of the constant a and the function $b(\cdot)$ as small as possible.

Due to the very general nonlinear structure we wish to accommodate for, the conditions imposed on the function f_{x2} in Assumption 2(d) are on the whole somewhat involved. The validity of these conditions can still be straightforwardly checked for many GARCH and ACD models, as our subsequent examples show. Often one can also use the following simple lemma (whose proof is omitted) to verify Assumptions 2(d_1), (d_2), and (d_3).

Lemma 2 *Suppose the function f_{x2} has range $\mathbb{R}_+$ (instead of $\mathbb{R}$) and that, for all $x > 0$, the function $f_{x2}(\cdot, x)$ is (1) surjective, (2) continuous, and (3) both monotonically increasing and continuously differentiable on (l, ∞), where $l \geq 0$. Then Assumptions 2(d_1), (d_2), and (d_3) hold.*

Assumption 2(d) enables us to prove our results without knowing anything about, not even existence of, the conditional density of X_t given $X_{t-1} = x$. In previous proofs of geometric ergodicity it has been quite typical to make explicit use of this conditional density and its properties (cf., e.g., Lu (1998) and Lanne and Saikkonen (2005)). While often straightforward this approach can sometimes be rather awkward to use and then our general conditions can be very convenient.

As indicated in the Introduction, our assumptions cover nonlinear specifications not covered by related previous studies. For instance, Carrasco and Chen (2002) and Straumann and Mikosch (2006) both give conditions for the ergodicity of rather general classes of GARCH models which, however, do not contain some of the nonlinear models to be

discussed below. In particular, some smooth transition GARCH models (see Example 4 below) in which f_{x1} and/or f_{x2} in (11) are nonlinear functions of X_{t-1} cannot be handled in the framework of Carrasco and Chen (2002). The framework of Straumann and Mikosch (2006) is similar to ours in that nonlinear functions of a very general form are employed but these functions are restricted by a Lipschitz condition. Even though this condition works well in several cases it rules out models covered by our Assumption 2. For instance, discontinuous models such as threshold-type GARCH models (see Example 3 below) are ruled out (some of them are covered by Carrasco and Chen (2002)) and the same is true for some smooth transition GARCH models.

As discussed in the context of the more general model (4)–(5), X_t can be viewed as a separate Markov chain generated by (11) and with the associated transition probability measure defining the counterpart of $P_X(\cdot,\cdot)$ in Assumption 1. The following theorem shows that, from this perspective, X_t is V–geometrically ergodic, as required for its counterpart in Proposition 1.

Theorem 1 *Consider X_t as a Markov chain generated by (11) and, in addition to Assumption 2, suppose that $E\left[(a+b(\varepsilon_t))^k\right] < 1$. Then X_t is V_X–geometrically ergodic with $V_X(x) = 1 + x^k$.*

The moment condition assumed in Theorem 1 is convenient in the proof and it also enables us to obtain explicit results about existence of moments. However, if one is only interested in proving geometric ergodicity an alternative condition, based on the following result, can be employed.

Lemma 3 *Suppose that (1) $E\left[b(\varepsilon_t)^k\right] < \infty$ for some $k \in \mathbb{R}_+$ and (2) $E\left[\ln(a+b(\varepsilon_t))\right] < 0$. Then there exists a $k_0 \in (0,k)$ such that $E\left[(a+b(\varepsilon_t))^{k_0}\right] < 1$.*

The result of this lemma can be justified by following the arguments in Remark 2.9 of Basrak, Davis, and Mikosch (2002). Condition (1) was already used in Assumption 2(e) whereas condition (2) is an analog of the necessary and sufficient condition for the stationarity of the conventional GARCH(1,1) process obtained by Nelson (1990) and extended to a class of GARCH(1,1) models by Francq and Zakoïan (2006, Theorem 2)

and Meitz (2006). If condition (2) is used in Theorem 1 instead of $E\left[(a+b(\varepsilon_t))^k\right] < 1$ we can conclude the $\left(1+x^{k_0}\right)$–geometric ergodicity of X_t for some $k_0 \in (0,k)$ but the precise value of k_0 cannot be determined. An analog of the result of Lemma 3 is also employed by Francq and Zakoïan (2006, Theorem 3) to prove the geometric ergodicity of a class of GARCH(1,1) models (see the beginning of step (iii) in the proof of their Lemma 2). Their proof also assumes an analog of condition (1) of Lemma 3 which is not required to obtain stationarity (see the aforementioned references). As a final remark of Lemma 3 we note that condition (2) is implied by $E\left[(a+b(\varepsilon_t))^k\right] < 1$ $(k>0)$, as Jensen's inequality shows.

To be able to apply Proposition 1 and obtain useful results for the joint process (Y_t, X_t), concrete assumptions about the functions f_{y1} and f_{y2} are needed. In most applications of GARCH or ACD models, the function f_{y2} is assumed to be a power function, that is, $f_{y2}(x) = x^{1/d}$ for some positive real number d. This is also the assumption we will make. In the context of ACD models, the function f_{y1} is always assumed to be 0. In GARCH models, the most common specification for the 'in-mean part' has also been a power function, say $f_{y1}(x) = \mu_0 + \mu_1 x^{1/e}$ for some positive real number e. We will assume slightly less, only dominance by such a function. The following theorem gives an ergodicity result for the joint process (Y_t, X_t) for these cases.

Theorem 2 *Suppose that the assumptions of Theorem 1 are satisfied and that* $f_{y2}(x) = x^{1/d}$ *where* $d \in \mathbb{R}_+$. ***(a)*** *If* $f_{y1}(x) = 0$ *and* $E[|\varepsilon_t|^{dk}] < \infty$, *where* k *is as in Assumption 2, then* Z_t *is* V_Z*–geometrically ergodic with* $V_Z(y,x) = 1 + |y|^{dk} + x^k$. ***(b)*** *If* $|f_{y1}(x)| \leq \mu_0 + \mu_1 x^{1/e}$, *where* $e \in \mathbb{R}_+$, μ_0, $\mu_1 \geq 0$, $E[|\varepsilon_t|^{\min\{d,e\}k}] < \infty$, *and* k *is as in Assumption 2, then* Z_t *is* V_Z*–geometrically ergodic with* $V_Z(y,x) = 1 + |y|^{\min\{d,e\}k} + x^k$.

As already noted after Proposition 1, the V_X–geometric ergodicity of X_t implies the V–geometric ergodicity of Z_t with $V(y,x) = V_X(x)$. A drawback of this choice of the function V is that nothing can be concluded about the moments of Y_t. The usefulness of being able to use a more general choice of the function V becomes clear in Theorem 2 where results on the existence of moments of the stationary distribution of Y_t are also obtained. If the conditions of Lemma 3 are used instead of the moment condition of Theorem 1, the conclusions of Theorem 2 hold with an unknown $k_0 \in (0,k)$. In this case,

part (b) of the theorem establishes geometric ergodicity in GARCH–M models under mild moment conditions.

Our next result applies Proposition 2 and provides conditions which guarantee that the joint process (Y_t, X_t) is β–mixing.

Theorem 3 *Suppose that the assumptions of Theorems 1 and 2(a) (or 2(b)) are satisfied, and that Z_t is initialized from $Z_0 = (Y_0, X_0)$ with a distribution μ such that $E_\mu[V_X(F_x(X_0, Y_0))] < \infty$, where $F_x(\cdot, \cdot)$ denotes the function on the right hand side of (10). Then Z_t is β–mixing with geometrically decaying mixing numbers.*

In summary, Theorems 1–3 establish the V–geometric ergodicity and β–mixing for the general model defined by equations (9) and (10) or equation (11). Existence of moments also readily follows from the stated conditions (in particular, $\sup_{t\geq 1} E_\mu[|Y_t|^{dk}] < \infty$ (or $\sup_{t\geq 1} E_\mu[|Y_t|^{\min\{d,e\}k}] < \infty$), and $\sup_{t\geq 1} E_\mu[X_t^k] < \infty$, and these moments converge to the ones taken with respect to the stationary distribution π_Z at a geometric rate).

Concrete examples where Theorems 1–3 apply are discussed now. Depending on which formulation has been more common in the literature, the structure of each model is described by using either equation (10) or equation (11). For convenience, all the examples are summarized in Tables 1 and 2 where choices of the relevant functions and constants assumed in the preceding results are also provided. Because typical choices of the function f_{y1} were already discussed $f_{y1}(\cdot) = 0$ is here assumed, so only GARCH and ACD models are considered. The validity of Assumption 2 can be rather straightforwardly verified for most of the considered models (see, however, some remarks in Examples 3 and 4 below). The form the condition $E\left[(a + b(\varepsilon_t))^k\right] < 1$ of Theorem 1 takes in these cases is also displayed in Table 1 with $k = 1$. The parameter restrictions implied by this condition agree in each case with the corresponding conditions reported in earlier literature. The weaker log-moment condition of Lemma 3 sufficing for geometric ergodicity is also given in Table 1.

Example 1 (The GARCH–family of Hentschel (1995) and the ACD–family of Fernandes and Grammig (2006)). Consider the family of GARCH models of

Hentschel (1995), which can be written as (see eq. (A.2) and (A.3) of Hentschel (1995))

$$\begin{aligned} Y_t &= \sigma_t \varepsilon_t \\ \sigma_t^\lambda &= \omega + (\alpha\lambda f^\nu(\varepsilon_{t-1}) + \beta)\sigma_{t-1}^\lambda \\ f(\varepsilon_{t-1}) &= |\varepsilon_{t-1} - b| - c(\varepsilon_{t-1} - b), \end{aligned} \tag{12}$$

where we assume that $b \in \mathbb{R}$, $|c| \leq 1$, and the remaining parameters take positive values.[2] Defining $X_t = \sigma_t^\lambda$ we arrive at a formulation written in the form of (9) and (11) as $Y_t = X_t^{1/\lambda}\varepsilon_t$ and $X_t = \omega + \alpha\lambda X_{t-1} f^\nu(\varepsilon_{t-1}) + \beta X_{t-1}$. In addition to the conventional linear GARCH model, this family also nests several other popular GARCH models (see Hentschel (1995) for a list). For brevity, the abbreviation BC–GARCH model is used in Tables 1 and 2 (here BC is due to the Box-Cox transformation). Fernandes and Grammig (2006) consider a family of ACD models analogous to Hentschel's family of GARCH models. This family can be defined with exactly the same equations (12).

Applying the result of Theorem 1 with $k = 1$ to the conventional linear GARCH model gives the condition $\alpha + \beta < 1$ for $(1 + x)$–geometric ergodicity of the conditional variance process. This agrees with the necessary and sufficient condition for second order stationarity (of Y_t) obtained by Bollerslev (1986). If we only assume that $k > 0$ and $E\left[\ln\left(\beta + \alpha\varepsilon_t^2\right)\right] < 0$, Lemma 3 and Theorem 1 give the $\left(1 + x^{k_0}\right)$–geometric ergodicity for some $0 < k_0 < k$. Here the condition $E\left[\ln\left(\beta + \alpha\varepsilon_t^2\right)\right] < 0$ agrees with the necessary and sufficient condition for the (strict) stationarity and geometric ergodicity of the conditional variance process obtained in Nelson (1990) and Francq and Zakoïan (2006, Theorem 3 and Remark 3), respectively. More generally, one can similarly conclude that in the general model (12) the condition $E\left[\ln\left(\beta + \alpha\lambda f^\nu(\varepsilon_{t-1})\right)\right] < 0$ is necessary and sufficient for geometric ergodicity (the necessity can be demonstrated as in Francq and Zakoïan (2006, second part of Theorem 2)). Thus, even though our model accommodates very general forms of nonlinearity we can obtain good results in this special case. In particular, the moment conditions we need are mild and comparable to those in Francq and Zakoïan (2006) although our assumptions about the distribution of the innovation process ε_t are stronger than assumed in that paper.

[2]Hentschel (1995) also considers a slightly different formulation which includes the case $\lambda = 0$. We do not discuss this case.

Example 2 (The integrated GARCH model). Choosing $f^{\nu}(\varepsilon_{t-1}) = \varepsilon_{t-1}^2$, $\alpha\lambda = 1-\beta$, and $\lambda = 2$ shows that the IGARCH model is a special case of the BC–GARCH model. In the case of the IGARCH model, we also assume that $E[\varepsilon_t^2] = 1$ regardless of the value of k, an assumption also made by Nelson (1990). As in that paper one can then see that condition (2) of Lemma 3, that is, $E[\ln(\beta + (1-\beta)\varepsilon_t^2)] < 0$, is automatically satisfied. Thus, in the IGARCH model the conditional variance process is $\left(1 + x^{k_0}\right)$–geometrically ergodic with some positive k_0. Using Theorem 2 we can even say more about the value of k_0. Unlike in all other models we have to assume that the condition $E\left[(a + b(\varepsilon_t))^k\right] < 1$ only holds for $k < 1$. Then, because $E[a + b(\varepsilon_t)] = E[\beta + (1-\beta)\varepsilon_t^2] = 1$, strict concavity and Jensen's inequality give $E[(a + b(\varepsilon_t))^k] < (E[a + b(\varepsilon_t)])^k = 1$ for $0 < k < 1$. Thus, for the IGARCH model Theorem 2 applies with $k < 1$ and $d = 2$ implying that the above mentioned $\left(1 + x^{k_0}\right)$–geometric ergodicity holds for any $k_0 < 1$. Consequently, Y_t has finite moments of orders smaller than 2. This is consistent with the well-known fact that the IGARCH process (that is Y_t) has a strictly stationary but not a second order stationary solution (see Nelson (1990)). Previously, properties of the IGARCH process were also studied by Ding and Granger (1996) who demonstrated its short memory nature by showing that an 'approximate' autocorrelation function of Y_t^2 decays to zero at a geometric rate. Our Theorem 3 makes this point more rigorous by showing that the process Y_t is β–mixing with geometrically decaying mixing numbers.

Example 3 (Threshold models). A popular non-linear GARCH model is the GJR–GARCH model of Glosten, Jaganathan, and Runkle (1993), where the equation corresponding to (10) is $X_t = \omega + (\alpha + \alpha^* \mathbf{1}(Y_{t-1} > 0))Y_{t-1}^2 + \beta X_{t-1}$. Here $\mathbf{1}(\cdot)$ is the indicator function. The GJR–GARCH model is an example of threshold GARCH (or TGARCH) models studied more generally by Ling (1999, Theorems 4.1 and 4.2) and Lee and Shin (2005, Example 5). In a simple first order version of this model, the equation corresponding to (10) is given by

$$X_t = \omega_j + \alpha_j Y_{t-1}^2 + \beta_j X_{t-1}, \quad \text{if} \quad Y_{t-1} \in [r_{j-1}, r_j),\ j = 1, \ldots, J, \tag{13}$$

where $-\infty = r_0 < r_1 < \cdots < r_J = \infty$ are the threshold values, and the remaining parameters satisfy $\omega_j > 0$, $\alpha_j > 0$, and $\beta_j \geq 0$. Replacing Y_{t-1}^2 with Y_{t-1} and $r_0 = -\infty$

with $r_0 = 0$ in this definition one obtains the threshold ACD (or TACD) model of Zhang, Russell, and Tsay (2001).

Verifying the validity of Assumption 2 for TGARCH and TACD models is more involved than for the preceding models. Details are therefore presented in the Appendix. Here we only mention that it is convenient to express the model in a form in which the counterpart of the function f_{x2} can take negative values. Note also that in this case the expression $a + b(\varepsilon_t)$ appearing in Theorem 1 and Lemma 3 takes a somewhat complicated form (see the Appendix). In Table 1 we therefore report parameter restrictions which imply the validity of the condition $E\left[(a + b(\varepsilon_t))^k\right] < 1$ of Theorem 1 and the condition $E\left[\ln(a + b(\varepsilon_t))\right] < 0$ of Lemma 3.

For the TGARCH model our results improve on those in Ling (1999, Theorems 4.1 and 4.2) and Lee and Shin (2005, Example 5) where only the existence of a stationary solution of the model is established but the uniqueness and ergodicity of this solution are not obtained. Although our framework is quite general and not exclusively designed for threshold models, the parameter restriction obtained from our Theorem 1 still agrees with the conditions given in these previous papers when $J = 2$, and, when $J > 2$, it even provides an improvement. In these papers it is also assumed that the chain (Y_t, X_t) is weak Feller (see Meyn and Tweedie (1993, p. 127)), a continuity assumption not satisfied by all TGARCH models and not required by our results. For the TACD model the condition obtained from Theorem 1 coincides with the sufficient condition for geometric ergodicity obtained by Zhang, Russell, and Tsay (2001). Finally, our Lemma 3 provides sufficient conditions for geometric ergodicity that are strictly weaker than the conditions reported both for the TGARCH and the TACD model in the aforementioned earlier papers.[3]

Example 4 (Smooth transition models). Smooth transition GARCH models were introduced by Hagerud (1996) and González-Rivera (1998), and discussed by Lundbergh and Teräsvirta (2002) and Lanne and Saikkonen (2005), while the ACD analogs were introduced by Meitz and Teräsvirta (2006). To obtain a general framework covering all these models, let G_1 and G_2 be functions with range $[0, 1]$, and $\omega > 0$, $\alpha > 0$, $\beta > 0$,

[3]For the TGARCH model an even better condition has recently been obtained by Cline (2006). However, this condition is quite complicated and obtaining an explicit condition in terms of the model parameters appears difficult, and therefore simulation or numerical methods may be necessary for its application.

$\omega^{**} > 0$, $\omega + \omega^* > 0$, $\alpha + \alpha^* > 0$, and $\beta + \beta^* > 0$. In the GARCH variant, the equation corresponding to (10) takes the form

$$X_t = \omega + \alpha Y_{t-1}^2 + \beta X_{t-1} + (\omega^* + \alpha^* Y_{t-1}^2) G_1(Y_{t-1}) + (\omega^{**} + \beta^* X_{t-1})\, G_2(X_{t-1}).$$

The ACD variant is otherwise similar except that on the right hand side Y_{t-1}^2 is twice replaced by Y_{t-1}.[4] For Assumption 2(d_2) to be satisfied we need to assume that the function G_1 is continuous from the left (or from the right). This, however, is not restrictive, because in practice G_1 is usually continuous. A sufficient condition for Assumption 2(d_3) to hold is that for large values of y the function $G_1(y)$ is differentiable and $G_1'(y) = o\,(y^{-2})$ as $y \to \infty$. This condition is satisfied in the two typical examples where G_1 is the cumulative distribution function of the logistic distribution or normal distribution. For the function G_2 much less needs to be assumed. For convenience, we may assume that the limit $\lim_{x\to\infty} G_2(x)$ exists, in which case the constant a in Table 2 has the stated form. Note that to satisfy the Lipschitz condition required in Proposition 3.1 of Straumann and Mikosch (2006), rather complicated restrictions need to be imposed on the model parameters and on the functions G_1 and G_2. These restrictions are also quite stringent. For instance, models in which the function G_2 has a continuous derivative attaining a value large enough at some point are ruled out. Such models can still satisfy our Assumption 2. Conditions for geometric ergodicity obtained from Theorem 1 and Lemma 3 are reported in Table 1, and the latter are an improvement compared to previously reported conditions.

As indicated earlier, the validity of Assumption 2(d) is relatively straightforward to check even for rather complicated nonlinear models. At least for some of the models discussed above alternative approaches, which require deriving the conditional density of X_t given $X_{t-1} = x$ and checking that it has suitable properties, can be cumbersome. This may be the case, for instance, if one has a smooth transition GARCH model with the function G_1 not monotonically increasing.

[4]Note that this model is not general enough to cover the recent smooth transition GARCH and ACD models of Medeiros and Veiga (2006) and Fernandes, Medeiros, and Veiga (2006). Because these models are nested in (9) and (10) they could be analyzed in our framework but at the cost of making the example more complex.

4 Conclusion

In this paper we have studied a general Markov model which contains an observable and an unobservable or hidden component. We gave conditions under which the V–geometric ergodicity of the hidden component viewed as a Markov chain of its own is inherited by the joint process formed of the two components. Conditions for β–mixing and existence of moments for the joint process were also obtained.

Results obtained for our general Markov model were applied to a wide class of models which includes as special cases many first order GARCH, GARCH–M, and ACD models with possibly complicated nonlinear structures. In some special cases our conditions for V–geometric ergodicity were seen to agree with conditions previously shown to be necessary and sufficient for stationarity and geometric ergodicity. For highly nonlinear models, our conditions provided an improvement over previous results, whereas in the case of GARCH–M models our results are, to the best of our knowledge, the first ones available. As our emphasis was on allowing for nonlinearities, we only considered first order models, which are also often found adequate in practice. Due to the very general nature of the employed assumptions, the results obtained for these models should be straightforward to apply. Compared to previous counterparts they appear especially convenient for models such as smooth transition GARCH models or their ACD versions where highly nonlinear structures have been considered. Extensions of our results to general higher-order GARCH, GARCH–M, and ACD models forms an interesting topic for future research.

	Model equations	Condition in Theorem 1 with $k=1$ (Condition (2) in Lemma 3)
GARCH	$X_t = \omega + \alpha Y_{t-1}^2 + \beta X_{t-1}$	$\alpha+\beta<1$ ($E[\ln(\beta+\alpha\varepsilon_t^2)]<0$)
ACD	$X_t = \omega + \alpha Y_{t-1} + \beta X_{t-1}$	$\alpha+\beta<1$ ($E[\ln(\beta+\alpha\varepsilon_t)]<0$)
GJR–GARCH	$X_t = \omega + (\alpha+\alpha^*\mathbf{1}(Y_{t-1}>0))Y_{t-1}^2 + \beta X_{t-1}$	$\alpha+\alpha^*/2+\beta<1$ [(1)] ($E[\ln(\beta+(\alpha+\alpha^*\mathbf{1}(\varepsilon_t>0))\varepsilon_t^2)]<0$)
TGARCH	$X_t = \sum_{j=1}^J(\omega_j+\alpha_j Y_{t-1}^2+\beta_j X_{t-1})\mathbf{1}(Y_{t-1}\in[r_{j-1},r_j))$, where $-\infty=r_0<r_1<\cdots<r_J=\infty$	$\max\{\alpha_1,\alpha_J\}+\max\{\beta_1,\beta_J\}<1$ [(2)] ($E[\ln(\max\{\beta_1,\beta_J\}+\max\{\alpha_1,\alpha_J\}\varepsilon_t^2)]<0$ [(3)])
TACD	$X_t = \sum_{j=1}^J(\omega_j+\alpha_j Y_{t-1}+\beta_j X_{t-1})\mathbf{1}(Y_{t-1}\in[r_{j-1},r_j))$, where $0=r_0<r_1<\cdots<r_J=\infty$	$\alpha_J+\beta_J<1$ [(2)] ($E[\ln(\beta_J+\alpha_J\varepsilon_t)]<0$ [(3)])
ST–GARCH	$X_t = \omega+\alpha Y_{t-1}^2+\beta X_{t-1}$ $+(\omega^*+\alpha^* Y_{t-1}^2)G_1(Y_{t-1})+(\omega^{**}+\beta^* X_{t-1})G_2(X_{t-1})$	$\alpha+\max\{\alpha^*,0\}+\beta+\beta^*G_2(\infty)<1$ [(4)] ($E[\ln(\beta+\beta^*G_2(\infty)+(\alpha+\max\{\alpha^*,0\})\varepsilon_t^2)]<0$) [(4)]
ST–ACD	$X_t = \omega+\alpha Y_{t-1}+\beta X_{t-1}$ $+(\omega^*+\alpha^* Y_{t-1})G_1(Y_{t-1})+(\omega^{**}+\beta^* X_{t-1})G_2(X_{t-1})$	$\alpha+\max\{\alpha^*,0\}+\beta+\beta^*G_2(\infty)<1$ [(4)] ($E[\ln(\beta+\beta^*G_2(\infty)+(\alpha+\max\{\alpha^*,0\})\varepsilon_t)]<0$) [(4)]
BC–GARCH	$X_t = \omega+\beta X_{t-1}+\alpha\lambda X_{t-1}f^\nu(\varepsilon_{t-1})$, and $Y_t = X_t^{1/\lambda}\varepsilon_t$	$E[\beta+\alpha\lambda f^\nu(\varepsilon_t)]<1$ ($E[\ln(\beta+\alpha\lambda f^\nu(\varepsilon_t))]<0$)
BC–ACD	$X_t = \omega+\beta X_{t-1}+\alpha\lambda X_{t-1}f^\nu(\varepsilon_{t-1})$, and $Y_t = X_t^{1/\lambda}\varepsilon_t$	$E[\beta+\alpha\lambda f^\nu(\varepsilon_t)]<1$ ($E[\ln(\beta+\alpha\lambda f^\nu(\varepsilon_t))]<0$)

Table 1: Summary of the discussed examples: Model equations, the form of condition $E[(a+b(\varepsilon_t))^k]<1$ in Theorem 1 with $k=1$, and the form of condition $E[\ln(a+b(\varepsilon_t))]<0$ in Lemma 3 (see Table 2 for the definitions of a and $b(\cdot)$). Notes: [(1)] Assuming ε_t has a symmetric distribution. [(2)] This is a condition implying the validity of $E[(a+b(\varepsilon_t))]<1$. [(3)] This is a condition implying the validity of $E[(\max\{\beta_1,\beta_J\}+\max\{\alpha_1,\alpha_J\}\varepsilon_t^2)^{k_0}]<1$ (TGARCH) or $E[(\beta_J+\alpha_J\varepsilon_t)^{k_0}]<1$ (TACD), and hence of $E[(a+b(\varepsilon_t))^{k_0}]<1$, for some $k_0>0$. [(4)] $G_2(\infty)$ is used as a shorthand notation for $\lim_{x\to\infty}G_2(x)$.

	$f_{y2}(x)$	$f_{x1}(x)$	$f_{x2}(f_{y2}(x)\varepsilon, x)$ [(1)]	a	$b(\varepsilon)$	c
GARCH	$x^{1/2}$	$\omega + \beta x$	αy^2	β	$\alpha\varepsilon^2$	0
ACD	x	$\omega + \beta x$	αy	β	$\alpha\varepsilon$	0
GJR–GARCH	$x^{1/2}$	$\omega + \beta x$	$(\alpha + \alpha^* \mathbf{1}(y > 0))y^2$	β	$(\alpha + \alpha^* \mathbf{1}(\varepsilon > 0))\varepsilon^2$	0
TGARCH [(2)]	$x^{1/2}$	$\beta_* x$	$\sum_{j=1}^{J} \{(\omega_j + \alpha_j y^2 + \beta_j x)$ $\mathbf{1}(y \in [r_{j-1}, r_j))\} - \beta_* x$	β_*	$\alpha_* \varepsilon^2$ $+ \max \beta_j \mathbf{1}(\varepsilon^2 < r_*^2/M)$ [(3)]	$\max \omega_j + r_*^2 \max \alpha_j$ $+M \max \beta_j$ [(3)]
TACD	x	$\beta_J x$	$\sum_{j=1}^{J} \{(\omega_j + \alpha_j y + \beta_j x)$ $\mathbf{1}(y \in [r_{j-1}, r_j))\} - \beta_J x$	β_J	$\alpha_J \varepsilon$ $+ \max \beta_j \mathbf{1}(\varepsilon < r_{J-1}/M)$ [(3)]	$\max \omega_j + r_{J-1} \max \alpha_j$ $+M \max \beta_j$ [(3)]
ST–GARCH	$x^{1/2}$	$\omega + \beta x$ $+(\omega^{**} + \beta^* x)G_2(x)$	$\omega^* G_1(y)$ $+(\alpha + \alpha^* G_1(y))y^2$	β $+\beta^* G_2(\infty)$ [(4)]	$(\alpha + \max\{\alpha^*, 0\})\varepsilon^2$	$\lvert\omega^*\rvert$
ST–ACD	x	$\omega + \beta x$ $+(\omega^{**} + \beta^* x)G_2(x)$	$\omega^* G_1(y)$ $+(\alpha + \alpha^* G_1(y))y$	β $+\beta^* G_2(\infty)$ [(4)]	$(\alpha + \max\{\alpha^*, 0\})\varepsilon$	$\lvert\omega^*\rvert$
BC–GARCH	$x^{1/\lambda}$	$\omega + \beta x$	$\alpha\lambda x f^\nu(\varepsilon)$	β	$\alpha\lambda f^\nu(\varepsilon)$	0
BC–ACD	$x^{1/\lambda}$	$\omega + \beta x$	$\alpha\lambda x f^\nu(\varepsilon)$	β	$\alpha\lambda f^\nu(\varepsilon)$	0

Table 2: Summary of the discussed examples (continued): Choices of the relevant functions and constants. The function f_{y1} is omitted as in every case $f_{y1}(x) = 0$. Notes: [(1)] Two different formulations, corresponding to equations (10) and (11), are used to achieve notational convenience. [(2)] For TGARCH we denote $\alpha_* = \max\{\alpha_1, \alpha_J\}$, $\beta_* = \max\{\beta_1, \beta_J\}$, and $r_* = \max\{|r_1|, |r_{J-1}|\}$ (as in the Appendix). [(3)] M is a (large) constant chosen in the Appendix. [(4)] $G_2(\infty)$ is used as a shorthand notation for $\lim_{x\to\infty} G_2(x)$.

Appendix: Proofs

Validity of Assumption 1 for the model (4)–(5). First conclude from (4) that the conditional probability distribution of Y_t given $X_t = x$ is defined by

$$\pi_{Y|X}(A \mid x) = \int \mathbf{1}(F_y(x,\zeta) \in A)\, P_\zeta(d\zeta), \qquad A \in \mathcal{B}(\mathcal{Y}),$$

where $\mathbf{1}(\cdot)$ is the indicator function and $P_\zeta(\cdot)$ signifies the probability distribution of ζ_t. Similarly, the transition probability measure of the Markov chain X_t defined by (6) is

$$P_X(x, A) = \int \mathbf{1}(G_x(x,\zeta) \in A)\, P_\zeta(d\zeta), \qquad A \in \mathcal{B}(\mathcal{X}),$$

from which the corresponding n–step transition probability measure can be derived (cf. Meyn and Tweedie (1993, p. 78)). To derive the probability measure $\widetilde{P}_X^n(\cdot,\cdot)$, let $z = (y, x)$ be an initial value and set $\widetilde{z} = \widetilde{z}(z) = F_x(z)$. Then conclude from equation (5) that $X_1 = \widetilde{z}$ and $X_2 = F_x(\widetilde{z}, F_y(\widetilde{z}, \zeta_1))$. Interpreting $\widetilde{z} = \widetilde{z}(z)$ as a function of z one obtains $\widetilde{P}_X^2(z,\cdot)$ from this whereas equation (6) yields $X_2 = G_x(\widetilde{z}, \zeta_1)$, $\widetilde{z} \in \mathcal{X}$. Thus, we get

$$\begin{aligned} \widetilde{P}_X^2(z, A) &= \int \mathbf{1}(G_x(\widetilde{z}(z), \zeta) \in A)\, P_\zeta(d\zeta) \\ &= \int \mathbf{1}(G_x(\widetilde{z}, \zeta) \in A)\, P_\zeta(d\zeta) = P_X(\widetilde{z}, A), \qquad A \in \mathcal{B}(\mathcal{X}). \end{aligned}$$

By induction it can be seen that $\widetilde{P}_X^n(z, A) = P_X^{n-1}(\widetilde{z}, A)$ for all $n \geq 2$ and, because the validity of equation (2) is straightforward to check, it follows that Assumption 1 applies to the model defined by (4) and (5) with $j = 1$ and the function λ given by $\lambda = F_x$. ■

Proof of Proposition 1. The proof is based on ideas similar to those in the proof of Proposition 4 of Carrasco and Chen (2002). Set $\pi_Z(\cdot) = \pi_{Y|X}(\cdot \mid x)\, \pi_X(\cdot)$ where $\pi_X(\cdot)$ signifies the stationary probability measure related to a Markov chain with transition probability measure $P_X(\cdot,\cdot)$. First note that

$$\begin{aligned} \int_Z V_Z(z)\pi_Z(dz) &= \int_{\mathcal{X}} \pi_X(dx) \int_{\mathcal{Y}} V_Z(y,x)\pi_{Y|X}(dy|x) \qquad (14) \\ &\leq \int_{\mathcal{X}} \pi_X(dx) c V_X(x) \\ &< \infty, \end{aligned}$$

where we have used the assumed condition $\int_{\mathcal{Y}} V_Z(y,x)\pi_{Y|X}(dy|x) \leq cV_X(x)$, for all $x \in \mathcal{X}$, and the V_X–geometric ergodicity of X_t. Then, for every $z_0 = (y_0, x_0) \in \mathcal{Y} \times \mathcal{X}$ and $n > j$,

$$\begin{aligned} & \sup_{s:|s|\leq V_Z} \left| \int_{\mathcal{Y}\times\mathcal{X}} \left[P_Z^n\left(z_0, dz\right) - \pi_Z\left(dz\right)\right] s\left(z\right) \right| \\ = & \sup_{s:|s|\leq V_Z} \left| \int_{\mathcal{X}} \left[\widetilde{P}_X^n\left(z_0, dx\right) - \pi_X\left(dx\right)\right] \left(\int_{\mathcal{Y}} \pi_{Y|X}\left(dy \mid x\right) s\left(y,x\right) \right) \right| \qquad (15) \\ \leq & \; c \sup_{v:|v|\leq V_X} \left| \int_{\mathcal{X}} \left[P_X^{n-j}\left(\widetilde{x}_0, dx\right) - \pi_X\left(dx\right)\right] v\left(x\right) \right|, \end{aligned}$$

where $\widetilde{x}_0 = \widetilde{x}\left(z_0\right) = \lambda(z_0)$. Here the equality follows from (2) and the definition of $\pi_Z(\cdot)$. In the inequality we have used Assumption 1(b) and the fact that, for any function s with $|s(\cdot)| \leq V_Z(\cdot)$,

$$\begin{aligned} \left| \int_{\mathcal{Y}} \pi_{Y|X}\left(dy|x\right) s\left(y,x\right) \right| & \leq \int_{\mathcal{Y}} \pi_{Y|X}\left(dy \mid x\right) \left|s\left(y,x\right)\right| \\ & \leq \int_{\mathcal{Y}} \pi_{Y|X}\left(dy \mid x\right) V_Z\left(y,x\right) \\ & \leq cV_X(x). \end{aligned}$$

Because $P_X\left(\cdot,\cdot\right)$ is assumed to be V_X–geometrically ergodic the last quantity in (15) can be bounded by a term of the form $\varrho^n M_{\widetilde{x}_0}$, where $\varrho < 1$ and $M_{\widetilde{x}_0} < \infty$. Thus, the same is true for the first quantity, implying that Z_t is V_Z–geometrically ergodic. ■

Proof of Proposition 2. By Proposition 2.4 of Liebscher (2005), Z_t is β–mixing with geometrically decaying mixing numbers if (i) $E_\mu[V_X(\lambda(X_0, Y_0))] < \infty$ and (ii) Z_t is Q–geometrically ergodic in the sense of Liebscher (2005) with $Q(z) = V_X(\lambda(x,y))$. Condition (i) holds by assumption (b). For condition (ii), we first need to show that $E_{\pi_Z}[V_X(\lambda(X_t, Y_t))] < \infty$. This is obtained from (14) by replacing $V_Z(z)$ with $V_X(\lambda(x,y))$ and using assumption (c) in conjunction with the V_X-geometric ergodicity of X_t. As for the remaining part of condition (ii), notice that from (15) and (8) we find that

$$\begin{aligned} \sup_{s:|s|\leq V_Z} \left| \int_{\mathcal{Y}\times\mathcal{X}} [P_Z^n(z_0, dz) - \pi_Z(dz)]s(z) \right| & \leq c \sup_{v:|v|\leq V_X} \left| \int_{\mathcal{X}} [P_X^{n-j}(\tilde{x}_0, dx) - \pi_X(dx)]v(x) \right| \\ & \leq \rho^n R V_X(\tilde{x}_0) \\ & = \rho^n R V_X(\lambda(x_0, y_0)) \end{aligned}$$

for some $\rho \in (0,1)$ and $R < \infty$ (here ρ^{-j} has been absorbed into R). Considering functions $s\left(\cdot\right) \leq 1$ completes the proof of condition (ii) (see the definition of Q–geometric ergodicity in Liebscher (2005, p. 671)). ■

Proof of Lemma 1. Consider the model (4)–(5) and suppose the assumptions of Proposition 2 are satisfied apart from (c). Recall that now Assumption 1 holds with the function λ given by $\lambda = F_x$. This in conjunction with the definition of the conditional probability distribution $\pi_{Y|X}(\cdot \mid x)$ and equations (4), (5), and (6) shows that

$$\begin{aligned}\int_{\mathcal{Y}} \pi_{Y|X}(dy \mid x)\, V_X(\lambda(x,y)) &= E[V_X(F_x(X_t, Y_t)) \mid X_t = x] \\ &= E\left[V_X\left(G_x(x, \zeta_t)\right)\right] \\ &= E\left[V_X\left(X_{t+1}\right) \mid X_t = x\right] \\ &= \int_{\mathcal{X}} P_X(x, dw)\, V_X(w).\end{aligned}$$

For simplicity, set $\int_{\mathcal{X}} \pi_X(dw) V_X(w) = C$ and note that $C < \infty$ by the assumed V_X-geometric ergodicity of $P_X(\cdot,\cdot)$. Thus, using (8) with $n = 1$ we find that, for all $x \in \mathcal{X}$,

$$\begin{aligned}\left|\int_{\mathcal{X}} P_X(x, dw)\, V_X(w)\right| &\leq \left|\int_{\mathcal{X}} P_X(x, dw)\, V_X(w) - \int_{\mathcal{X}} \pi_X(dw) V_X(w)\right| + C \\ &\leq (\varrho R + C)\, V_X(x),\end{aligned}$$

where $\varrho R + C < \infty$. Combining the preceding inequalities yields the stated result. ■

Proof of Theorem 1. We use $\mu_{Leb}(\cdot)$ to signify the Lebesgue measure on $\mathbb{R}$ and $P_X(\cdot,\cdot)$ the transition probability measure obtained when X_t is viewed as a separate Markov chain generated by (11). Due to the imposed assumptions, the state space of X_t is $\mathcal{X} = [\underline{f}, \infty)$. The proof consists of showing that X_t is irreducible and aperiodic, that an appropriate small set exists, and that the so-called drift condition is satisfied with the function V_X (for definitions of these concepts, see Meyn and Tweedie (1993)). Irreducibility, the existence of a small set, and aperiodicity are first proven in Lemmas 4, 5, and 6, respectively.

Lemma 4 *If the assumptions of Theorem 1 hold there exist real numbers $\underline{l}$ and $\overline{l}$ such that $(\underline{l}, \overline{l}) \subset \mathcal{X}$ and the Markov chain X_t is φ–irreducible with $\varphi(\cdot) = \mu_{Leb}(\cdot \cap (\underline{l}, \overline{l}))$.*

Proof. By assumption $E\left[(a + b(\varepsilon_t))^k\right] < 1$. Therefore we can choose an $\epsilon > 0$ such that

$$E\left[(a + \epsilon + b(\varepsilon_t))^k\right] < 1. \tag{16}$$

By Assumptions 2(c) and (e) we can now choose an $M_\epsilon \in \mathbb{R}_+$ such that

$$f_{x2}(f_{y2}(x)\varepsilon_t, x) \leq x b(\varepsilon_t) + \frac{1}{2}\epsilon x \tag{17}$$

and

$$f_{x1}(x) \leq ax + \frac{1}{2}\epsilon x \tag{18}$$

for $x \in \mathcal{X}$ and $x > M_\epsilon$. Define the sets $S_{1\epsilon} = \{x \in \mathcal{X} : x > M_\epsilon\}$ and $S_{2\epsilon} = \{x \in \mathcal{X} : x \leq M_\epsilon\}$. Without loss of generality M_ϵ can be chosen large enough that $S_{2\epsilon}$ is nonempty. Clearly $\mathcal{X} = S_{1\epsilon} \cup S_{2\epsilon}$. From (16) it follows that $a + \frac{1}{2}\epsilon < 1$, and hence we can without loss of generality also assume that $M_\epsilon > (1 - a - \frac{1}{2}\epsilon)^{-1} \inf I$, where I denotes the interval in Assumption 2(d_1) (this fact will be used later in the proof of Lemma 6).

We shall next prove the following four results:

I) $\forall x \in S_{1\epsilon} : \exists n \in \mathbb{Z}_+ : P^n(x, S_{2\epsilon}) > 0$

II) $\forall x \in S_{2\epsilon} : P(x, A \cap (\underline{l}, \overline{l})) > 0$ whenever $\mu_{Leb}(A \cap (\underline{l}, \overline{l})) > 0$

III) $\inf_{x \in S_{2\epsilon}} P(x, A \cap (\underline{l}, \overline{l})) > 0$ whenever $\mu_{Leb}(A \cap (\underline{l}, \overline{l})) > 0$

IV) $\forall x \in S_{1\epsilon} : \exists n \in \mathbb{Z}_+ : P^{n+1}(x, A \cap (\underline{l}, \overline{l})) > 0$ whenever $\mu_{Leb}(A \cap (\underline{l}, \overline{l})) > 0$

Establishing II and IV will complete the proof of Lemma 4, while III will be used later in the proof of Lemma 5.

Proof of I. Let $t \in \mathbb{Z}_+$ and suppose that $X_{t-1} \in S_{1\epsilon}$. Using (17) and (18) we find that $X_t \leq X_{t-1}\,(a + \epsilon + b(\varepsilon_{t-1}))$ and, since both sides are positive, $X_t^k \leq X_{t-1}^k\,(a + \epsilon + b(\varepsilon_{t-1}))^k$. Next consider the event

$$\Omega_n = \left\{ (a + \epsilon + b(\varepsilon_{t-1}))^k < E\left[(a + \epsilon + b(\varepsilon_{t-1}))^k\right], t = 1, \ldots, n \right\},$$

where n is a positive integer. The nonconstancy and continuity of $b(\cdot)$ on some open set implies that the probability of Ω_n is positive for every n. Thus, on the event Ω_n,

$$X_t^k \leq X_{t-1}^k \cdot E\left[(a + \epsilon + b(\varepsilon_{t-1}))^k\right], \tag{19}$$

where by (16) the expectation is < 1.

Now choose an arbitrary $x \in S_{1\epsilon}$, and denote $X_0 = x$. Using (19) inductively we have, for arbitrary $n \in \mathbb{Z}_+$ and on the event Ω_n, that

$$X_n^k \leq x^k \cdot \left\{ E\left[(a + \epsilon + b(\varepsilon_{t-1}))^k\right] \right\}^n \tag{20}$$

as long as $X_1, \ldots, X_{n-1} \in S_{1\epsilon}$. Since $E[(a+\epsilon+b(\varepsilon_{t-1}))^k] < 1$, the right-hand-side of (20) will eventually be less than or equal to M_ϵ^k when n is chosen large enough, and for such n we will have $X_n \in S_{2\epsilon}$. Since the probability of the event Ω_n is positive for every n, we have thus completed the proof of I.

Proof of II and III. Since the functions f_{x1} and f_{x2} are bounded on bounded subsets of their domain there exist positive and finite real numbers M_1 and M_2 such that

$$\sup_{x \in S_{2\epsilon}} f_{x1}(x) \leq M_1 \quad \text{and} \quad \sup_{x \in S_{2\epsilon}} f_{x2}(R,x) \leq M_2. \tag{21}$$

Define $\underline{l} = \max\{M_1 + M_2, M_\epsilon + 1\}$, and choose an arbitrary $\overline{l} > \underline{l}$ (note that the fact $\underline{l} > M_\epsilon$ is used only later in the proof of Lemma 6).

Now choose an arbitrary set A such that $\mu_{Leb}(A \cap (\underline{l}, \overline{l})) > 0$. Furthermore, choose an arbitrary $x \in S_{2\epsilon}$. For the 1–step transition probability from x to A it holds that

$$\begin{aligned} P(x,A) &= \int_{-\infty}^{\infty} \mathbf{1}\left(f_{x1}(x) + f_{x2}\left(f_{y2}(x)\,\varepsilon, x\right) \in A\right) \phi_\varepsilon(\varepsilon)\, d\varepsilon \\ &\geq \int_{R/f_{y2}(x)}^{\infty} \mathbf{1}\left(f_{x1}(x) + f_{x2}\left(f_{y2}(x)\,\varepsilon, x\right) \in A\right) \phi_\varepsilon(\varepsilon)\, d\varepsilon. \end{aligned}$$

According to Assumption 2(d) $f_{x1}(x) + f_{x2}\left(f_{y2}(x)\,\varepsilon, x\right)$ is monotonically increasing with respect to ε on the integration range, and thus, making a transformation of variables $v = f_{x1}(x) + f_{x2}\left(f_{y2}(x)\,\varepsilon, x\right)$, we have

$$\begin{aligned} P(x,A) &\geq \int_{\{v > f_{x2}(R,x) + f_{x1}(x)\}} \mathbf{1}(v \in A)\, \phi_\varepsilon\left(\frac{f_{x2}^{-1}(v - f_{x1}(x), x)}{f_{y2}(x)}\right) \\ &\qquad \times \frac{1}{f_{y2}(x)} \frac{\partial f_{x2}^{-1}(v - f_{x1}(x), x)}{\partial v} dv \\ &\geq \int_{A \cap (\underline{l}, \overline{l})} \phi_\varepsilon\left(\frac{f_{x2}^{-1}(v - f_{x1}(x), x)}{f_{y2}(x)}\right) \frac{1}{f_{y2}(x)} \frac{\partial f_{x2}^{-1}(v - f_{x1}(x), x)}{\partial v} dv. \end{aligned}$$

The boundedness conditions for f_{x1}, f_{x2}, f_{y2}, $\partial f_{x2}^{-1}(v,x)/\partial v$, and ϕ_ε imply that

$$\inf_{x \in S_{2\epsilon}, v \in A \cap (\underline{l}, \overline{l})} \phi_\varepsilon\left(\frac{f_{x2}^{-1}(v - f_{x1}(x), x)}{f_{y2}(x)}\right) \frac{1}{f_{y2}(x)} \frac{\partial f_{x2}^{-1}(v - f_{x1}(x), x)}{\partial v} \geq \epsilon_*$$

for some positive ϵ_*, and therefore $P(x,A) \geq \epsilon_* \mu_{Leb}(A \cap (\underline{l}, \overline{l}))$. Because the set A can clearly be replaced by $A \cap (\underline{l}, \overline{l})$ both II and III are obtained from

$$\inf_{x \in S_{2\epsilon}} P(x, A \cap (\underline{l}, \overline{l})) \geq \epsilon_* \mu_{Leb}(A \cap (\underline{l}, \overline{l})). \tag{22}$$

Proof of IV. Choose an arbitrary set A such that $\mu_{Leb}(A \cap (\underline{l}, \overline{l})) > 0$, and an arbitrary $x \in S_{1\epsilon}$. According to I, we can choose an integer n such that $P^n(x, S_{2\epsilon}) > 0$. Now, by the Chapman-Kolmogorov equation (Meyn and Tweedie (1993, Theorem 3.4.2, p. 67))

$$\begin{aligned} P^{n+1}(x, A \cap (\underline{l}, \overline{l})) &= \int_{\mathcal{X}} P^n(x, dy) P(y, A \cap (\underline{l}, \overline{l})) \\ &\geq \int_{S_{2\epsilon}} P^n(x, dy) P(y, A \cap (\underline{l}, \overline{l})) \\ &\geq \int_{S_{2\epsilon}} P^n(x, dy) \epsilon_* \mu_{Leb}(A \cap (\underline{l}, \overline{l})) \\ &= P^n(x, S_{2\epsilon}) \epsilon_* \mu_{Leb}(A \cap (\underline{l}, \overline{l})) \\ &> 0, \end{aligned}$$

where the first inequality follows from the fact that $S_{2\epsilon} \subset \mathcal{X}$, and the second inequality follows from (22). This completes the proof of IV. ■

Lemma 5 *If the assumptions of Theorem 1 hold the set $S_{2\epsilon}$ is small.*

Proof. Equation (22) shows that equation (5.14) of Meyn and Tweedie (1993) holds with the measure $\epsilon_* \mu_{Leb}(\cdot \cap (\underline{l}, \overline{l}))$. Thus, the set $S_{2\epsilon}$ is small by the definition of a small set. ■

Lemma 6 *If the assumptions of Theorem 1 hold the Markov chain X_t is aperiodic.*

Proof. By Proposition A1.1 of Chan (1990), the aperiodicity of X_t obtains if

$$\forall x \in A : \big(P(x, A) > 0 \text{ and } P^2(x, A) > 0\big) \tag{23}$$

for some small set A such that $\varphi(A) > 0$. We shall show this holds with the set $(\underline{l}, \overline{l})$.

To this end, it suffices to prove that for all open subsets A of $S_{1\epsilon}$ and for every $x \in A$, $P(x, A) > 0$ and $P^2(x, A) > 0$. Let A be an arbitrary open subset of $S_{1\epsilon}$ and $x \in A$ be arbitrary. Because $x > M_\epsilon$, we have by (18) and (16), that $f_{x1}(x) \leq \big(a + \frac{1}{2}\epsilon\big) x < x$, and furthermore that $x - f_{x1}(x) \geq x - \big(a + \frac{1}{2}\epsilon\big) x = x(1 - a - \frac{1}{2}\epsilon)$. As $M_\epsilon > (1 - a - \frac{1}{2}\epsilon)^{-1} \inf I$, we therefore have $x - f_{x1}(x) > \inf I$, where I again denotes the interval in Assumption 2(d_1). The same assumption now implies that there exists a u such that $f_{x1}(x) + f_{x2}(u, x) = x$. Hence we can also find an $\underline{e} \in (\underline{\varepsilon}, \infty)$ such that $f_{x1}(x) + f_{x2}(f_{y2}(x)\, \underline{e}, x) = x$. Since the set A is open, we can choose a $\delta > 0$ such that $(x - \delta, x + \delta) \subset A$, and the continuity

from the right (alternatively, continuity from the left) of $f_{x2}(\cdot, x)$ ensures that for a such δ, there exists an $\overline{e} > \underline{e}$ (alternatively, $\overline{e} < \underline{e}$) such that

$$\varepsilon \in (\underline{e}, \overline{e}) \Rightarrow f_{x1}(x) + f_{x2}(f_{y2}(x)\,\varepsilon, x) \in (x - \delta, x + \delta) \tag{24}$$

(alternatively, $\varepsilon \in (\overline{e}, \underline{e})$). Thus, we can conclude that

$$\begin{aligned} P(x, A) &\geq P(x, (x - \delta, x + \delta)) \\ &= \Pr(f_{x1}(x) + f_{x2}(f_{y2}(x)\,\varepsilon_t, x) \in (x - \delta, x + \delta)) \\ &\geq \Pr(\varepsilon_t \in (\underline{e}, \overline{e})) \\ &> 0, \end{aligned}$$

where the second inequality follows from (24) and the third from the assumed positivity of $\phi_\varepsilon(\cdot)$. In addition, by the Chapman-Kolmogorov equation,

$$P^2(x, A) = \int_{\mathcal{X}} P(x, dy) P(y, A) \geq \int_{(x-\delta, x+\delta)} P(x, dy) P(y, A) > 0.$$

Hence the assertion made is proven. Since this holds, in particular, for the set $(\underline{l}, \overline{l})$ the condition (23) is established with $A = (\underline{l}, \overline{l})$.

By Lemma 4, $\varphi\left((\underline{l}, \overline{l})\right) > 0$. To establish that the set $(\underline{l}, \overline{l})$ is small consider first the proof of parts II and III in Lemma 4 but with the set $S_{2\epsilon}$ replaced by $(\underline{l}, \overline{l})$. Repeating the arguments in that proof we can find an $\epsilon'_* > 0$ and an open interval $(\underline{l}', \overline{l}')$ such that the transition probabilities from $(\underline{l}, \overline{l})$ to $(\underline{l}', \overline{l}')$ are positive and

$$\inf_{x \in (\underline{l}, \overline{l})} P(x, A \cap (\underline{l}', \overline{l}')) \geq \epsilon'_* \mu_{Leb}(A \cap (\underline{l}', \overline{l}')) > 0 \tag{25}$$

whenever $\mu_{Leb}(A \cap (\underline{l}', \overline{l}')) > 0$. Equation (5.14) of Meyn and Tweedie (1993) is now satisfied with the measure $\epsilon'_* \mu_{Leb}(\cdot \cap (\underline{l}', \overline{l}'))$, and thus the set $(\underline{l}, \overline{l})$ is small. ■

Finishing the proof of Theorem 1.

Given Lemmas 4, 5, and 6, it now suffices to show that condition (15.3) of Meyn and Tweedie (1993) holds with the function $V_X(x) = 1 + x^k$. This in turn is the case if there exist constants $c_1 > 0$ and $c_2 < \infty$ such that

$$E[V_X(X_t) \mid X_{t-1} = x] \leq (1 - c_1) V_X(x) + c_2 \mathbf{1}(x \in S_{2\epsilon}) \quad \text{for all } x \in \mathcal{X}. \tag{26}$$

The expectation in (26) can be written as

$$E\left[1 + X_t^k \mid X_{t-1} = x\right] = 1 + E\left[\left(f_{x1}(x) + f_{x2}(f_{y2}(x)\varepsilon_{t-1}, x)\right)^k\right].$$

Suppose first that $x \in S_{1\epsilon}$. As in the proof of part I of Lemma 4 we have

$$\begin{aligned} 1 + E\left[\left(f_{x1}(x) + f_{x2}(f_{y2}(x)\varepsilon_{t-1}, x)\right)^k\right] \leq & 1 + x^k E\left[\left(a + \epsilon + b(\varepsilon_{t-1})\right)^k\right] \\ = & \left(1 - \frac{x^k\left(1 - E\left[\left(a + \epsilon + b(\varepsilon_{t-1})\right)^k\right]\right)}{1 + x^k}\right)(1 + x^k). \end{aligned}$$

Redefining M_ϵ if necessary we can without loss of generality assume that $M_\epsilon > 1$. Then $x > 1$ and $x^k/(1 + x^k) > 1/2$. Since $E\left[\left(a + \epsilon + b(\varepsilon_{t-1})\right)^k\right] < 1$, it follows that

$$\begin{aligned} & 1 + E\left[\left(f_{x1}(x) + f_{x2}(f_{y2}(x)\varepsilon_{t-1}, x)\right)^k\right] \\ < & \left(1 - \frac{1}{2}\left(1 - E\left[\left(a + \epsilon + b(\varepsilon_{t-1})\right)^k\right]\right)\right)(1 + x^k). \end{aligned}$$

Defining $c_1 = \frac{1}{2}\left(1 - E\left[\left(a + \epsilon + b(\varepsilon_{t-1})\right)^k\right]\right)$ shows that (26) holds for all $x \in S_{1\epsilon}$.

Suppose now that $x \in S_{2\epsilon}$. Then, by the first inequality in (21) and Assumption 2(e),

$$\begin{aligned} 1 + E\left[\left(f_{x1}(x) + f_{x2}(f_{y2}(x)\varepsilon_{t-1}, x)\right)^k\right] & \leq 1 + E\left[\left(M_1 + c + xb(\varepsilon_{t-1})\right)^k\right] \\ & \leq 1 + E\left[\left(M_1 + c + M_\epsilon b(\varepsilon_{t-1})\right)^k\right] \\ & < \infty. \end{aligned}$$

Defining $c_2 = 1 + E\left[\left(M_1 + c + M_\epsilon b(\varepsilon_{t-1})\right)^k\right]$ and noting that $(1 - c_1)V(x)$ is always positive shows that (26) holds also for all $x \in S_{2\epsilon}$. Since $\mathcal{X} = S_{1\epsilon} \cup S_{2\epsilon}$, this completes the proof of V_X–geometric ergodicity. ■

Proof of Theorem 2. The fact that the Markov chain $Z_t = (Y_t, X_t)$ satisfies Assumption 1 follows from the discussion after this assumption. Also, X_t viewed as a separate Markov chain is V_X–geometrically ergodic by Theorem 1. Hence, by Proposition 1, it remains to be proven that $\int_{\mathcal{Y}} \pi_{Y|X}(dy|x)V_Z(y, x) \leq cV_X(x)$ for all $x \in \mathcal{X}$ and some $c < \infty$.

The conditional probability distribution of Y_t given $X_t = x$ is

$$\pi_{Y|X}(dy|x) = \frac{1}{f_{y2}(x)}\phi_\varepsilon\left(\frac{y - f_{y1}(x)}{f_{y2}(x)}\right)dy.$$

Thus, since $V_X(x) = 1 + x^k$, part (a) follows by observing that

$$\int_{\mathcal{Y}} \pi_{Y|X}(dy|x) V_Z(y,x) = 1 + x^k + E\left[|x^{1/d}\varepsilon_t|^{dk}\right] \leq (1 + x^k)(1 + E\left[|\varepsilon_t|^{dk}\right]).$$

Consider now part (b), and suppose first that $d \leq e$. Similarly as above,

$$\begin{aligned}\int_{\mathcal{Y}} \pi_{Y|X}(dy|x) V_Z(y,x) &= 1 + x^k + E\left[|x^{1/d}\varepsilon_t + f_{y1}(x)|^{dk}\right] \\ &\leq 1 + x^k + C_1\left(E\left[|x^{1/d}\varepsilon_t|^{dk}\right] + |f_{y1}(x)|^{dk}\right) \\ &\leq 1 + x^k + C_1\left(x^k E\left[|\varepsilon_t|^{dk}\right] + C_2(\mu_0^{dk} + \mu_1^{dk} x^{dk/e})\right) \qquad (27)\end{aligned}$$

for some constants C_1 and C_2, where we have twice used Loève's c_r–inequality (see Davidson (1994, p. 140)). In (27), $x^{dk/e} \leq \max\{1, x^k\} \leq (1 + x^k)$, and hence the expression in (27) is smaller than $C_3(1 + x^k)$ for some constant C_3. The case $d > e$ can be proven in an analogous way. ■

Proof of Theorem 3. It was established in the proof of Theorem 2 that under current assumptions the Markov chain $Z_t = (Y_t, X_t)$ satisfies the conditions of Proposition 1. The validity of condition (a) of Proposition 2 follows from Theorem 15.0.1 of Meyn and Tweedie (1993), because we have established the validity of their condition (15.3) in the proof of Theorem 1 (see equation (26)). Condition (b) is satisfied by assumption because in the present case $\lambda = F_x$ whereas condition (c) is redundant by Lemma 1. The results follow by applying Proposition 2 with the functions V_Z from Theorem 2. ■

Validity of Assumption 2 for TGARCH and TACD models. Consider the TGARCH model, which, denoting $R_j = [r_{j-1}, r_j)$, can be written as

$$X_t = \sum_{j=1}^{J} \left(\omega_j + \alpha_j Y_{t-1}^2 + \beta_j X_{t-1}\right) \mathbf{1}\left(Y_{t-1} \in R_j\right).$$

For simplicity, denote $\alpha_* = \max\{\alpha_1, \alpha_J\}$, $\beta_* = \max\{\beta_1, \beta_J\}$, and $r_* = \max\{|r_1|, |r_{J-1}|\}$. Defining $f_{x1}(x) = \beta_* x$ and $f_{x2}(x^{1/2}\varepsilon, x) = \sum_{j=1}^{J} (\omega_j + \alpha_j x\varepsilon^2 + \beta_j x)\, \mathbf{1}\left(x^{1/2}\varepsilon \in R_j\right) - \beta_* x$ we have $X_t = f_{x1}(X_{t-1}) + f_{x2}(X_{t-1}^{1/2}\varepsilon_{t-1}, X_{t-1})$. The validity of conditions (b), (c), (d_2), and (d_3) of Assumption 2 is rather clear. For condition (d_1) it suffices to note that $[\omega_J + \alpha_J r_{J-1}, \infty) \subseteq f_{x2}([r_{J-1}, \infty), x) = [\omega_J + \alpha_J r_{J-1} + (\beta_J - \beta_*)\, x, \infty)$ for all x.

For condition (e), note that $\sum_{j=2}^{J-1}(\alpha_j x\varepsilon^2)\,\mathbf{1}\left(x^{1/2}\varepsilon\in R_j\right)\leq r_*^2\max\alpha_j$, $\sum_{j=1}^{J}\omega_j\mathbf{1}\left(x^{1/2}\varepsilon\in R_j\right)\leq\max\omega_j$, and for any positive M (which is to be chosen shortly)

$$\begin{aligned}\sum_{j=2}^{J-1}(\beta_j x)\,\mathbf{1}\left(x^{1/2}\varepsilon\in R_j\right) &\leq x\max\beta_j\mathbf{1}\left(x\varepsilon^2\leq r_*^2\right)\\ &= x\max\beta_j\mathbf{1}\left(x\varepsilon^2\leq r_*^2\right)\left[\mathbf{1}\left(x\leq M\right)+\mathbf{1}\left(x>M\right)\right]\\ &\leq M\max\beta_j+x\max\beta_j\mathbf{1}\left(x\varepsilon^2\leq r_*^2\right)\mathbf{1}\left(x>M\right)\\ &\leq M\max\beta_j+x\max\beta_j\mathbf{1}\left(\varepsilon^2\leq r_*^2/M\right).\end{aligned}$$

Therefore

$$\begin{aligned}f_{x2}(x^{1/2}\varepsilon,x) &= \sum_{j=1}^{J}\left(\omega_j+\alpha_j x\varepsilon^2+\beta_j x\right)\mathbf{1}\left(x^{1/2}\varepsilon\in R_j\right)-\beta_* x\\ &\leq \max\omega_j+r_*^2\max\alpha_j+M\max\beta_j+x\max\beta_j\mathbf{1}\left(\varepsilon^2\leq r_*^2/M\right)\\ &\quad+(\alpha_1 x\varepsilon^2+\beta_1 x)\mathbf{1}\left(x^{1/2}\varepsilon\in R_1\right)+(\alpha_J x\varepsilon^2+\beta_J x)\mathbf{1}\left(x^{1/2}\varepsilon\in R_J\right)-\beta_* x\\ &\leq \max\omega_j+r_*^2\max\alpha_j+M\max\beta_j+x\max\beta_j\mathbf{1}\left(\varepsilon^2\leq r_*^2/M\right)+\alpha_* x\varepsilon^2\end{aligned}$$

and, denoting $c=(\max\omega_j+r_*^2\max\alpha_j+M\max\beta_j)$ and $b(\varepsilon)=(\alpha_*\varepsilon^2+\max\beta_j\mathbf{1}(\varepsilon^2\leq r_*^2/M))$, we have established the validity of the inequality in condition (e).

It remains to be verified that the moment condition in (e) is satisfied. For this, we next establish that, for any $k>0$, if $E\left[(\alpha_*\varepsilon_t^2+\beta_*)^k\right]<1$, then $E[(a+b(\varepsilon_t))^k]<1$ (and thus also $E[b(\varepsilon_t)^k]<\infty$ in condition (e)). In addition to completing the verification of Assumption 2, this gives an easily verifiable condition which implies the validity of the moment restriction in Theorem 1. When $k\leq 1$, Loève's c_r–inequality gives $E[(a+b(\varepsilon_t))^k]\leq E\left[(\alpha_*\varepsilon_t^2+\beta_*)^k\right]+E\left[(\max\beta_j\mathbf{1}(\varepsilon_t^2\leq r_*^2/M))^k\right]=E\left[(\alpha_*\varepsilon_t^2+\beta_*)^k\right]+\max\beta_j^k E\left[\mathbf{1}(\varepsilon_t^2\leq r_*^2/M)\right]$. By choosing M sufficiently large, the last term can be made arbitrarily small, and hence $E[(a+b(\varepsilon_t))^k]<1$ for a suitable choice of M. Suppose now that $k>1$. By Minkowski's inequality we have $E[(a+b(\varepsilon_t))^k]\leq\{(E\left[(\alpha_*\varepsilon_t^2+\beta_*)^k\right])^{1/k}+(E\left[(\max\beta_j\mathbf{1}(\varepsilon_t^2\leq r_*^2/M))^k\right])^{1/k}\}^k$ and, by choosing M sufficiently large, the second expectation can again be made small enough so that $E[(a+b(\varepsilon_t))^k]<1$. This completes establishing the validity of the conditions for the TGARCH model. The required arguments for the TACD model are similar, although slightly simpler, and are omitted.

■

References

Basrak, B., R. A. Davis, and T. Mikosch (2002): "Regular variation of GARCH processes," *Stochastic Processes and their Applications*, 99, 95–115.

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 nonnegative time series," *Journal of Econometrics*, 52, 115–127.

Carrasco, M., and X. Chen (2002): "Mixing and moment properties of various GARCH and stochastic volatility models," *Econometric Theory*, 18, 17–39.

Carrasco, M., and X. Chen (2005): "Erratum: Mixing and moment properties of various GARCH and stochastic volatility models," unpublished manuscript.

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

Cline, D. B. H. (2006): "Stability of nonlinear stochastic recursions with application to nonlinear AR–GARCH models," unpublished manuscript.

Ding, Z., and C. W. J. Granger (1996): "Modeling volatility persistence of speculative returns: A new approach," *Journal of Econometrics*, 73, 185–215.

Davidson, J. (1994): *Stochastic Limit Theory*. Oxford University Press, Oxford.

Doukhan, P. (1994): *Mixing: Properties and Examples*. Springer-Verlag, New York.

Duan, J.-C. (1997): "Augmented GARCH(p,q) process and its diffusion limit," *Journal of Econometrics*, 79, 97–127.

Engle, R. F. (1982): "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation," *Econometrica*, 50, 987–1008.

Engle, R. F., and J. R. Russell (1998): "Autoregressive conditional duration: a new model for irregularly spaced transaction data," *Econometrica*, 66, 1127–1162.

Fernandes, M., and J. Grammig (2006): "A family of autoregressive conditional duration models," *Journal of Econometrics*, 130, 1–23.

Fernandes, M., M. C. Medeiros, and A. Veiga (2006): "A (semi-)parametric functional coefficient autoregressive conditional duration model," Department of Economics, PUC-Rio, Textos para discussão No. 535.

Francq, C., and J.-M. Zakoïan (2006): "Mixing properties of a general class of GARCH(1,1) models without moment assumptions," *Econometric Theory*, 22, 815–834.

Genon-Catalot, V., T. Jeantheau, and C. Larédo (2000): "Stochastic volatility models as hidden Markov models and statistical applications," *Bernoulli*, 6, 1051–1079.

Glosten, L. W., R. Jaganathan, and D. 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.

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

Hagerud, G. E. (1996): "A smooth transition ARCH model for asset returns," Working Paper Series in Economics and Finance No. 162, Stockholm School of Economics.

Hentschel, L. (1995): "All in the family: Nesting symmetric and asymmetric GARCH models," *Journal of Financial Economics*, 39, 71–104.

Kristensen, D. (2007): "Uniform ergodicity of a class of Markov chains with applications to time series models," unpublished manuscript.

Lanne, M., and P. Saikkonen (2005): "Nonlinear GARCH models for highly persistent volatility," *Econometrics Journal*, 8, 251–276.

Lee, O., and D. W. Shin (2005): "On stationarity and β–mixing property of certain nonlinear GARCH(p,q) models," *Statistics and Probability Letters*, 73, 25–35.

Liebscher, E. (2005): "Towards a unified approach for proving geometric ergodicity and mixing properties of nonlinear autoregressive processes," *Journal of Time Series Analysis*, 26, 669–689.

Ling, S. (1999): "On the probabilistic properties of a double threshold ARMA conditional heteroskedastic model," *Journal of Applied Probability*, 36, 688–705.

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

Ling, S., and M. McAleer (2003): "Asymptotic theory for a vector ARMA–GARCH model," *Econometric Theory*, 19, 280–310.

Lu, Z. (1998): "On the geometric ergodicity of a non-linear autoregressive model with an autoregressive conditional heteroscedastic 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 (2002): "Evaluating GARCH models," *Journal of Econometrics*, 110, 417–435.

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

Medeiros, M. C., and A. Veiga (2006): "Modeling multiple regimes in financial volatility with a flexible coefficient GARCH(1,1) model," Department of Economics, PUC-Rio, Textos para discussão No. 486.

Meitz, M. (2006): "A necessary and sufficient condition for the strict stationarity of a family of GARCH processes," *Econometric Theory*, 22, 985–988.

Meitz, M., and T. Teräsvirta (2006): "Evaluating models of autoregressive conditional duration," *Journal of Business and Economic Statistics*, 24, 104–124.

Meyn, S. P., and R. L. Tweedie (1993): *Markov Chains and Stochastic Stability*. Springer-Verlag, London.

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

Straumann, D., and T. Mikosch (2006): "Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: a stochastic recurrence equations approach," *Annals of Statistics*, 34, 2449–2495.

Zhang, M. Y., J. R. Russell, and R. S. Tsay (2001): "A nonlinear autoregressive conditional duration model with applications to financial transaction data," *Journal of Econometrics*, 104, 179–207.