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

Mika Meitz
Stockholm School of Economics

Pentti Saikkonen
University of Helsinki

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

October 7, 2004

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 (a form of) geometric ergodicity of the unobservable component is inherited by the joint process formed of the two components are given. This immediately implies the existence of initial values such that the joint process is strictly stationary and β–mixing. In addition to this, conditions for β–mixing and existence of moments for the joint process are also provided in the case of (possibly) nonstationary initial values. 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 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. Address correspondence to: Mika Meitz, Department of Economic Statistics, Stockholm School of Economics, P. O. Box 6501, SE–113 83 Stockholm, Sweden; e-mail: mika.meitz@hhs.se; or to: Pentti Saikkonen, Department of 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, mixing and has bounded 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. Conditions which imply β–mixing and existence of moments in the case of nonstationary initial values are also provided. Motivated by other potential applications, these results are first obtained for a general class of Markov models defined in terms of transition probability measures.

We also consider a sub-class of the aforementioned general class of Markov models and show that it contains many GARCH and ACD models as special cases. Results obtained for our general class of Markov models are applied to this sub-class. For simplicity, we concentrate on the leading case of first order GARCH and ACD models but, on the other hand, 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 (in press), 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 considered.

The approach used in this paper has previous counterparts. 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). Our general class of Markov models may therefore be seen as a counterpart of the generalized hidden Markov model of Carrasco and Chen (2002) which also applies to GARCH and ACD models.

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 previously appeared in Nelson (1990), Bougerol and Picard (1992), Zhang, Russell, and Tsay (2001), Ling and McAleer (2002), Ling and McAleer (2003), and Lanne and Saikkonen (2004). 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.

The rest of this paper is organized as follows. Our general class of Markov models is studied in Section 2 where the generalized hidden Markov model of Carrasco and Chen (2002) is also discussed. 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 Models

We shall first discuss the generalized hidden Markov model of Carrasco and Chen (2002) and demonstrate why it is not an appropriate model class for our purposes. 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 various other mixing concepts are discussed.

The model of Carrasco and Chen (2002) involves 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. The formal definition states that Y_t is a generalized hidden Markov model with hidden chain X_t if the following three conditions are satisfied.[1]

(i) X_t is an unobserved stationary Markov chain on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$.

(ii) 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.

(iii) The conditional distribution of Y_t given $X_t = x$ does not depend on t.

The hidden Markov model of Genon-Catalot, Jeantheau, and Larédo (2000) assumes conditions (i) and (iii) but replaces (ii) by the following condition.

(ii)' For all t, given $(X_0, \ldots, X_t)$, $Y_0, \ldots, Y_t$ are conditionally independent and the conditional distribution of Y_s only depends on X_s $(0 \leq s \leq t)$.

Proposition 4 of Carrasco and Chen (2002) attempts to extend previous similar results of Genon-Catalot, Jeantheau, and Larédo (2000) which show that a hidden Markov model is ergodic (and strong mixing) when the involved hidden Markov chain is ergodic (and strong mixing). Specifically, this proposition assumes conditions (i)–(iii) and contains two assertions. The first assertion claims that the joint process (Y_t, X_t) is a geometrically ergodic Markov chain if the hidden chain X_t is geometrically ergodic. According to the second assertion, Y_t is stationary β–mixing if X_t is stationary β–mixing and the decay of the mixing coefficients of Y_t is at least as fast as that of X_t. Unfortunately, however, the employed definition of generalized hidden Markov model does not guarantee the validity of these results, as the following example shows.

[1] In this paper, a conditional probability distribution refers to a regular conditional probability distribution the existence of which is assumed and, unless specified, a stationary process will henceforth mean a strictly stationary process.

Let ε_t be a sequence of n.i.d.$(0,1)$ random variables and consider the model

$$\begin{aligned} Y_t &= X_t + \zeta_t \\ X_t &= \xi_t, \end{aligned} \tag{1}$$

where $\zeta_t = \varepsilon_{2t}$ and $\xi_t = \varepsilon_t$, $t = 1, 2, \ldots$. The model is extended for $t = 0$ by assuming that ζ_0 and ξ_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. From the imposed assumptions it also follows that 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, given $X_t = x$, the conditional distribution of Y_t is $N(x, 1)$ for all t.[2] Conditions (i), (ii), and (iii) of the preceding definition of generalized hidden Markov model are thus satisfied. However, the results of Proposition 4 of Carrasco and Chen (2002) do not hold for this model. For instance, $\mathrm{Cov}(Y_1, Y_2) = \mathrm{Cov}(\varepsilon_2, \varepsilon_2) = 1$ and $\mathrm{Cov}(Y_3, Y_4) = 0$ and therefore Y_t is not stationary although X_t is. Also, for all $t \geq 1$, $\mathrm{Cov}(Y_t, Y_{2t}) = \mathrm{Cov}(\varepsilon_{2t}, \varepsilon_{2t}) = 1$, which implies that Y_t is not strong mixing and, hence, not β–mixing (cf. Proposition 1, p. 4, and Theorem 3, p. 9, in Doukhan (1994)). It is also clear that 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.

The problem with the preceding definition of generalized hidden Markov model seems to be that the defining conditions do not bind the joint distribution of the process (Y_t, X_t) enough for useful results to be obtainable. For instance, in the case of model (1) the processes ζ_t and ξ_t are individually well behaved normally distributed i.i.d. processes but jointly they are neither stationary nor mixing. In the hidden Markov model of Genon-Catalot, Jeantheau, and Larédo (2000) difficulties of this kind are avoided because condition (ii)' ensures sufficient structure for the distribution of the joint process (Y_t, X_t) and, therefore, undesirable cases such as (1) are excluded.[3]

For simplicity, denote $Z_t = (Y_t, X_t)$. One possibility to obtain sufficient structure for the joint process Z_t is to assume that it is a (time homogeneous) Markov chain on $(\mathcal{Z}, \mathcal{B}(\mathcal{Z}))$ where $\mathcal{Z} = \mathcal{Y} \times \mathcal{X}$ and $\mathcal{B}(\mathcal{Z}) = \mathcal{B}(\mathcal{Y} \times \mathcal{X})$.[4] As the example above shows, this assumption is not implied by conditions (i)–(iii) although it can be justified for the GARCH models studied by Carrasco and Chen (2002). For instance, consider the standard GARCH(1,1) model

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

where $\varepsilon_t \sim$ i.i.d.$(0,1)$ with ε_t independent of (u_s, h_s), $s < t$, and the parameters satisfy the usual conditions $\omega > 0$, $\alpha \geq 0$, $\beta \geq 0$, and $\alpha + \beta < 1$. Substituting h_t from the latter equation to the former shows that (u_t, h_t) can be viewed as 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 $h_t = \omega + \beta h_{t-1} + \alpha h_{t-1} \varepsilon_{t-1}^2$.

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 the n–step transition probability measure of the Markov chain Z_t and, assuming conditions (ii) and (iii), let $\pi_{Y|X}(\cdot \mid x)$ signify the conditional probability distribution of Y_t given $X_t = x$ ($P_Z^1(\cdot, \cdot) = P_Z(\cdot, \cdot)$ and similarly

[2] From the formulation of the model it is seen that $(X_t, Y_{t-1}, X_{t-1}, \ldots, Y_0, X_0)$ can be expressed as a function of X_t and the random variables $\zeta_0, \xi_0, \varepsilon_1, \ldots, \varepsilon_{t-1}, \varepsilon_{t+1}, .., \varepsilon_{2(t-1)}$. Because these random variables are independent of X_t and Y_t the stated results readily follow.

[3] Note that condition (ii)' is not satisfied by model (1) because, given $(X_0, \ldots, X_t)$, the conditional distribution of Y_s does not necessarily depend on X_s only. Because model (1) is a special case of Example 1 of Carrasco and Chen (2002) this also means that, contrary to the authors' statement, their example is not a hidden Markov model.

[4] Unless otherwise stated, Markov chains are always assumed to be time homogeneous in this paper.

for other transition probability measures). 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 condition (i) of their generalized hidden Markov model and replace the latter factor by $P_X^n(x, dx)$, the n–step transition probability measure of the Markov chain X_t. However, it is not clear how this replacement can be justified in GARCH models, for example. Although h_t in (2) can be treated as a Markov chain of its own this Markov chain is not identical to the generation process of h_t obtained from the joint process (u_t, h_t). Specifically, given the initial value (u_0, h_0), the joint process implies that $h_1 = \omega + \beta h_0 + \alpha u_0^2$ whereas $h_1 = \omega + \beta h_0 + \alpha h_0 \varepsilon_0^2$ is obtained when h_t is treated as a separate Markov chain. Thus, if the joint process (u_t, h_t) is viewed as a Markov chain the conditional probability distribution of h_1 also depends on the initial value u_0. This demonstrates that in the general case it is not clear that one can replace the initial value z by x in the conditional probability $\Pr(dx \mid Z_0 = z)$.

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

$$P_Z^n(z, dz) = \pi_{Y|X}(dy \mid x) \widetilde{P}_X^n(z, dx). \tag{3}$$

As far as potential generalizations of the hidden Markov model are concerned, 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 (2) and the related two-dimensional Markov chain. As noticed above, we have $h_1 = \omega + \beta h_0 + \alpha u_0^2$ and, as can be easily checked, $h_2 = \omega + \beta \widetilde{h}_0 + \alpha \widetilde{h}_0 \varepsilon_1^2$ where $\widetilde{h}_0 = \omega + \beta h_0 + \alpha u_0^2$. Thus, the generating mechanism of h_2 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 (2) as a pattern we now replace the probability measure $\widetilde{P}_X^n(z, \cdot)$ in (3) 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 (3) 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 $\widetilde{x} = \widetilde{x}(z) = \lambda(z)$.*

As the preceding discussion indicates, it is implicit in Assumption 1(a) that conditions (ii) and (iii) 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 some separate Markov chain on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ with suitably defined initial values. However, condition (i) is not satisfied (not even without the word 'stationary') because the transition probability measure of X_t depends on the initial value of the joint process (Y_t, X_t). Neither is condition (ii)' satisfied, as an example below shows. Importantly and contrary to conditions (i)–(iii), we also assume that the joint process (Y_t, X_t) is a Markov chain, a condition not implied by (i)–(iii). Note also that we do not assume stationarity. Assuming stationarity would simplify some of the subsequent derivations but at the cost of not obtaining results of practical relevance (see Proposition 2 and Theorem 3 below).

It is straightforward to check that Assumption 1 holds for the GARCH(1,1) model (2) 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. For a more general example, which contains the GARCH(1,1) model (2) as a special case, consider the model

$$\begin{aligned} Y_t &= F_y(X_t, \zeta_t) && (4) \\ X_t &= F_x(X_{t-1}, Y_{t-1}), && (5) \end{aligned}$$

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. Substituting the right hand side of equation (5) for X_t in (4) shows that the joint process (Y_t, X_t) is a Markov chain. From (4) and (5) it also follows that the random vectors ζ_t and X_t are independent and that the latter has the representation

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

Thus, X_t can be viewed as a Markov chain of its own and studied in isolation from Y_t.

To verify Assumption 1, 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 $P_X(\cdot,\cdot)$ obtained from the Markov chain (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 representation of X_2. On the other hand, using (6) yields $X_2 = G_x(\widetilde{z}, \zeta_1)$, $\widetilde{z} \in \mathcal{X}$, so that altogether we have

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

Proceeding inductively it can further be seen that $\widetilde{P}_X^n(z, A) = P_X^{n-1}(\widetilde{z}, A)$ holds for all $n \geq 2$. This means that $P_X(\cdot,\cdot)$ is the transition probability measure obtained from (6) by treating X_t as a separate Markov chain whereas $j = 1$ and the function λ is given by $\lambda = F_x$. Because the validity of equation (3) is also straightforward to check Assumption 1 applies to the model defined by (4) and (5).[5]

We shall now show that results, essentially the same as stated in Proposition 4 of Carrasco and Chen (2002), hold when the conditions therein are replaced by Assumption 1. We use the so-called V–geometric ergodicity of a Markov chain (see Meyn and Tweedie (1993, p. 356)). By definition, 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

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

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 $E\left(v\left(Z_n\right) \mid Z_0 = z\right)$.

The weakest form of V–geometric ergodicity obtains 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 $z \in \mathcal{Z}$. Because this convergence holds regardless of the initial value z a geometrically ergodic Markov chain can be thought of as exhibiting a form of 'stability'. 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 the Markov chain Z_t if the distribution of the initial value Z_0 is defined by the probability measure π_Z (see Meyn and Tweedie (1993, p. 230–231)).

A convenient feature of V–geometric ergodicity is that it implies existence of moments. For instance, if the Markov chain Z_t is initialized from the stationary distribution, existence of the expectation of $v\left(Z_t\right)$ for all v such that $|v(\cdot)| \leq V(\cdot)$ follows immediately from the definition of V–geometric ergodicity. Of course, it is also a direct consequence of (7) that, for such a v, the conditional expectation $E\left(v\left(Z_t\right) \mid Z_0 = z\right)$ converges at a geometric rate to the corresponding expectation taken with respect to the stationary distribution. Below results on existence of moments will also be established in the case of nonstationary initial values. However, first a result on V–geometric ergodicity of Z_t will be obtained.

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 $\int_{\mathcal{Y}} \pi_{Y|X}\left(dy \mid x\right) V_Z(y, x) \leq cV_X(x)$ 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 $V_Z\left(y, x\right) = V_X\left(x\right)$, although stronger and 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 Z_t is initialized from its stationary distribution, it further follows that Z_t is β–mixing with geometrically decaying mixing numbers (cf. Doukhan (1994, p. 4 and 89)). Thus, Proposition 1 provides us with results entirely similar to those stated in Proposition 4 of Carrasco

[5]This model is not a hidden Markov model which is obtained if equation (4) is retained but equation (5) is replaced by $X_t = F_x\left(X_{t-1}, \xi_t\right)$ where ξ_t is an i.i.d. sequence independent of the sequence ζ_t and of X_s, $s < t$. Substituting $F_x\left(X_{t-1}, \xi_t\right)$ for X_t in (4) then again gives the Markov chain representation of of (Y_t, X_t).

and Chen (2002). Using Proposition 1 one can therefore justify the proofs of Carrasco and Chen (2002) which rely on their Proposition 4.

Our next proposition extends the results on β–mixing and existence of moments discussed above to allow for nonstationary initial values. From a practical point of view, fixed initial values are of special interest because they are often used in simulation studies and estimation exercises involving, for example, GARCH and ACD models. From the viewpoint of developing asymptotic estimation and testing theory results on mixing and existence of moments are convenient for they justify the use of traditional limit theorems. 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 the following conditions hold.*

(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 $Z_0 = (Y_0, X_0)$.

(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, $\sup_{t\geq 1} E_\mu[v(Z_t)] < \infty$ for any function v such that $|v(\cdot)| \leq V_Z(\cdot)$, and these moments converge to the ones taken with respect to the stationary distribution π_Z at a geometric rate.

As already discussed, the results of this proposition are well-known when Z_t is initialized from its stationary distribution π_Z. Then Z_t is also stationary (see Meyn and Tweedie (1993, pp. 230–231)) and the conclusions of the proposition hold without conditions (a), (b) and (c), and for any function v such that $|v(\cdot)| \leq V_Z(\cdot)$ with V_Z as in Proposition 1. The extension to nonstationary initial values is based on recent results of Liebscher (2004). To be able to apply these results we need assumptions not needed in the case of stationary initial values. Our first additional assumption requires condition (8) which is slightly stronger than V_X-ergodicity of $P_X(\cdot,\cdot)$. This not restrictive, however, because condition (8) 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 many models. In particular, it is 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).*

It may also be noted that assumptions (a) and (c) 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 this condition is also necessary in our case.

3 GARCH and ACD models

In the same way as in Carrasco and Chen (2002), Propositions 1 and 2 of the previous section can be used to deduce ergodicity, mixing, and moment properties of GARCH models. In addition to various GARCH(1,1) models Carrasco and Chen (2002) also considered some higher-order GARCH models as well as examples of ACD models and autoregressive stochastic volatility models. For simplicity, we shall focus on the first order case but, on the other hand, allow for more complicated nonlinear structures than in previous work. 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.

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. Some concrete examples will be given later.

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 $f_{x1}(x) \le ax + o(x)$ 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 for* $x \geq \underline{f}$*.*

The requirement that the random variables ε_t have a density which is bounded away from zero on compact subsets of $(\underline{\varepsilon}, \infty)$ is satisfied in most situations of practical interest. 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 needed in the development of the subsequent theory, but they are hardly restrictive in practice. Especially the latter assumptions, though, suggest that some attention has 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 allow f_{x2} to also take negative values 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 (2004) 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 values 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. In a number of cases 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 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 (2004)). While often straightforward this approach can sometimes be rather awkward to use and then our general conditions can be very convenient.

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 $V_X(x) = 1 + x^k$.*

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 $|f_{y1}(x)| \le \mu_0 + \mu_1 x^{1/e}$, where $e \in \mathbb{R}_+$, μ_0, $\mu_1 \ge 0$, $\min\{d,e\}k \ge 1$, $E[|\varepsilon_t|^{\min\{d,e\}k}] < \infty$, 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 Y_t are also obtained. Our next result applies Proposition 2 and provides conditions which guarantee that the joint process (Y_t, X_t) is β–mixing and has finite moments.

Theorem 3 *Suppose that the assumptions of Theorems 1 and 2(i) (or 2(ii)) 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, $\sup_{t\ge1} E_\mu[|Y_t|^{dk}] < \infty$ (or $\sup_{t\ge1} E_\mu[|Y_t|^{\min\{d,e\}k}] < \infty$), and $\sup_{t\ge1} E_\mu[X_t^k] < \infty$. Furthermore, these moments converge to the ones taken with respect to the stationary distribution π_Z at a geometric rate.*

In summary, Theorems 1–3 establish the V–geometric ergodicity, β–mixing, and existence of moments for a general class of models which includes as special cases various GARCH, GARCH–M, and ACD models. Some concrete examples will be 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 that GARCH–M models are not included. The validity of Assumption 2 can be rather straightforwardly verified for most of the considered models (see, however, some notes concerning the threshold and smooth transition models 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.

Consider, for instance, 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.[6] Defining $X_t = \sigma_t^\lambda$ we arrive at a formulation written in the form of (9) and (11) as

$$\begin{aligned} Y_t &= X_t^{1/\lambda}\varepsilon_t \\ X_t &= \omega + \alpha\lambda X_{t-1} f^\nu(\varepsilon_{t-1}) + \beta X_{t-1}. \end{aligned}$$

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 used in the formulation of the model). Fernandes and Grammig (in press) 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) where, to use notation more conventional in the ACD–literature, σ might be replaced with ψ.

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. Our results also hold for the IGARCH model but, unlike for all other cases, with the condition $E\left[(a+b(\varepsilon_t))^k\right] < 1$ only holding for $k < 1$. To see this, notice that now $E[a + b(\varepsilon_t)] = E[\beta + (1-\beta)\,\varepsilon_t^2] = 1$ so that 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 3 should be applied with $k < 1$ and $d = 2$ implying that 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 results make this point more rigorous by showing that the process Y_t is β–mixing with geometrically decaying mixing numbers.

A popular non-linear GARCH model is the GJR–GARCH model of Glosten, Jaganathan, and Runkle (1993), where the equation corresponding to (10) is

$$X_t = \omega + (\alpha + \alpha^* \mathbf{1}\left(Y_{t-1} > 0\right))Y_{t-1}^2 + \beta X_{t-1}.$$

A restricted version of this model is nested in the family of Hentschel (1995). An ACD model resembling the GJR–GARCH model, and not nested by the family of ACD models of Fernandes and Grammig (in press), is the threshold ACD (or TACD) model of Zhang, Russell, and Tsay

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

(2001). In this model the parameter values are allowed to vary depending on a 'regime' determined by past values of Y_t. In a simple version of this model, the equation corresponding to (10) is

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

where J is the number of different regimes, $0 = r_0 < r_1 < \cdots < r_J = \infty$ are the threshold values, and the parameters satisfy $\omega_j > 0$, $\alpha_j \geq 0$, and $\beta_j \geq 0$. Verifying the validity of Assumption 2 for this model 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 moment condition $E\left[(a + b(\varepsilon_t))^k\right] < 1$ assumed in Theorem 1 takes a somewhat complicated form (see the Appendix). In Table 1 we therefore report a parameter restriction which implies the validity of this condition and also agrees with the sufficient condition for geometric ergodicity previously obtained by Zhang, Russell, and Tsay (2001).

Yet another class of non-linear GARCH and ACD models, not nested in the families of Hentschel (1995) or Fernandes and Grammig (in press), are the smooth transition GARCH and ACD models. The GARCH versions were introduced in Hagerud (1996) and González-Rivera (1998), and are also discussed in Lundbergh and Teräsvirta (2002) and Lanne and Saikkonen (2004), while the ACD analog was introduced in Meitz and Teräsvirta (2004). For these models, assume that G_1 and G_2 are functions with range $[0, 1]$, and that $\omega > 0$, $\alpha > 0$, $\beta > 0$, $\omega^{**} > 0$, $\omega + \omega^* > 0$, $\alpha + \alpha^* > 0$, and $\beta + \beta^* > 0$. In the GARCH variant, the dynamics of X_t corresponding to equation (10) are governed by

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

The ACD variant is otherwise similar except that on the right hand side Y_{t-1}^2 is twice replaced by Y_{t-1}. For Assumption 2(d_2) to be satisfied we need to assume that the function G_1 is either continuous from the left or continuous from the right. This, however, is not restrictive, because in practice G_1 is usually continuous or an indicator function of an interval with the latter possibility relevant for threshold type variants of the model. 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 $G_1'(y) = o\left(y^{-2}\right)$ as $y \to \infty$ (this condition ensures, in particular, that for large values of its first argument the function $f_{x2}(u, x)$ is monotonically increasing). This condition is satisfied when G_1 is the cumulative distribution function of a continuous random variable with sufficiently thin tails. Two typical examples are the cumulative distribution functions of the logistic distribution and 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.

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 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 provided. Unlike in many previous studies, these results were also shown to hold in the empirically relevant case where the initial values are nonstationary.

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. As our emphasis was on allowing for nonlinearities, we only considered the first order case, which is 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. For such models our results on β–mixing and existence of moments are also of importance for they open up the way to the development of the still missing asymptotic estimation and testing theory. This, as well as 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 $k=1$
GARCH	$X_t = \omega + \alpha Y_{t-1}^2 + \beta X_{t-1}$	$\alpha + \beta < 1$
ACD	$X_t = \omega + \alpha Y_{t-1} + \beta X_{t-1}$	$\alpha + \beta < 1$
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)
TACD	$X_t = \omega_j + \alpha_j Y_{t-1} + \beta_j X_{t-1}$	$\alpha_J + \beta_J < 1$ (2)
ST–GARCH	$X_t = \omega + \alpha Y_{t-1}^2 + \beta X_{t-1}$ $+(\omega^* + \alpha^* Y_{t-1}^2)G_1(Y_{t-1}) + (\omega^{**} + \beta^* X_{t-1})G_2(X_{t-1})$	$\alpha + \max\{\alpha^*, 0\} + \beta + \beta^* G_2(\infty) < 1$ (3)
ST–ACD	$X_t = \omega + \alpha Y_{t-1} + \beta X_{t-1}$ $+(\omega^* + \alpha^* Y_{t-1})G_1(Y_{t-1}) + (\omega^{**} + \beta^* X_{t-1})G_2(X_{t-1})$	$\alpha + \max\{\alpha^*, 0\} + \beta + \beta^* G_2(\infty) < 1$ (3)
BC–GARCH	$X_t = \omega + \beta X_{t-1} + \alpha\lambda X_{t-1} f^\nu(\varepsilon_{t-1})$, and $Y_t = X_t^{1/\lambda}\varepsilon_t$	$E[\beta + \alpha\lambda f^\nu(\varepsilon_t)] < 1$
BC–ACD	$X_t = \omega + \beta X_{t-1} + \alpha\lambda X_{t-1} f^\nu(\varepsilon_{t-1})$, and $Y_t = X_t^{1/\lambda}\varepsilon_t$	$E[\beta + \alpha\lambda f^\nu(\varepsilon_t)] < 1$

Table 1: Summary of the discussed examples: Model equations and the form the condition $E[(a+b(\varepsilon_t))^k] < 1$ of Theorem 1 takes with $k=1$. 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) $G_2(\infty)$ 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
TACD	x	$\beta_J x$	$\omega_j + \alpha_j y$	β_J	$\alpha_J \varepsilon$	$\max \omega_j + r_{J-1} \max \alpha_j$
			$+\beta_j x - \beta_J x$		$+ \max \beta_j \mathbf{1}(\varepsilon < r_{J-1}/M)$	$+M \max \omega_j$
ST–GARCH	$x^{1/2}$	$\omega + \beta x$	$\omega^* G_1(y)$	β	$(\alpha + \max\{\alpha^*, 0\})\varepsilon^2$	$\|\omega^*\|$
		$+(\omega^{**} + \beta^* x)G_2(x)$	$+(\alpha + \alpha^* G_1(y))y^2$	$+\beta^* G_2(\infty)$ [(2)]		
ST–ACD	x	$\omega + \beta x$	$\omega^* G_1(y)$	β	$(\alpha + \max\{\alpha^*, 0\})\varepsilon$	$\|\omega^*\|$
		$+(\omega^{**} + \beta^* x)G_2(x)$	$+(\alpha + \alpha^* G_1(y))y$	$+\beta^* G_2(\infty)$ [(2)]		
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)] $G_2(\infty)$ is used as a shorthand notation for $\lim_{x\to\infty} G_2(x)$.

Appendix: Proofs

Proof of Proposition 1. 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

$$\begin{aligned}\int_Z V_Z(z)\pi_Z(dz) &= \int_{\mathcal{Y}\times\mathcal{X}} V_Z(y,x)\pi_{Y|X}(dy|x)\pi_X(dx) \\ &= \int_{\mathcal{X}} \pi_X(dx)\int_{\mathcal{Y}} V_Z(y,x)\pi_{Y|X}(dy|x) \\ &\leq \int_{\mathcal{X}} \pi_X(dx) c V_X(x) \\ &< \infty,\end{aligned} \tag{13}$$

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 $n > j$,

$$\begin{aligned}&\sup_{s:|s|\leq V_Z}\left|\int_{\mathcal{Y}\times\mathcal{X}}\left[P_Z^n(z_0,dz) - \pi_Z(dz)\right]s(z)\right| \\ =\;& \sup_{s:|s|\leq V_Z}\left|\int_{\mathcal{X}}\left[\widetilde{P}_X^n(z_0,dx) - \pi_X(dx)\right]\left(\int_{\mathcal{Y}}\pi_{Y|X}(dy\mid x)\,s(y,x)\right)\right| \\ \leq\;& c\sup_{v:|v|\leq V_X}\left|\int_{\mathcal{X}}\left[P_X^{n-j}(\widetilde{x}_0,dx) - \pi_X(dx)\right]v(x)\right|,\end{aligned} \tag{14}$$

where $\widetilde{x}_0 = \widetilde{x}(z_0) = \lambda(z_0)$. Here the equality follows from (3) 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 $|s(\cdot)| \leq V_Z(\cdot)$,

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

Because $P_X(\cdot,\cdot)$ is assumed to be V_X–geometrically ergodic the last quantity in (14) 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. This completes the proof. ■

Proof of Proposition 2. By Proposition 2.4 of Liebscher (2004), Z_t is β–mixing with geometrically decaying mixing numbers if the following two conditions hold: (i) $E_\mu[V_X(\lambda(X_0,Y_0))] < \infty$, and (ii) Z_t is Q–geometrically ergodic in the sense of Liebscher (2004) with $Q(z) = V_X(\lambda(x,y))$. The first condition is satisfied by assumption (b). To verify condition (ii), we first need to show that $E_{\pi_Z}[V_X(\lambda(X_t,Y_t))] < \infty$. This, however, is obtained from the inequalities (13) 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.

To establish the remaining part of condition (ii) as well as the finiteness of the asserted moments, notice that from (14) and (8) we can conclude 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 $s(\cdot) \leq 1$ completes the proof of condition (ii) (see the definition of Q–geometric ergodicity in Liebscher (2004)). Furthermore, for any function s such that $|s(\cdot)| \leq V_Z(\cdot)$

$$\left| \int_{\mathcal{Z}} s(z) P_Z^n(z_0, dz) - \int_{\mathcal{Z}} s(z) \pi_Z(dz) \right| \leq \rho^n R V_X(\lambda(z_0))$$

and thus, integrating over the initial distribution μ,

$$\left| \int_{\mathcal{Z}_0} \int_{\mathcal{Z}} s(z) P_Z^n(z_0, dz) \mu(dz_0) - \int_{\mathcal{Z}_0} \int_{\mathcal{Z}} s(z) \pi_Z(dz) \mu(dz_0) \right| \leq \rho^n R \int_{\mathcal{Z}_0} V_X(\lambda(z_0)) \mu(dz_0).$$

Here $\mathcal{Z}_0 = \mathcal{Z}$ is used to clarify the order of integration and

$$\int_{\mathcal{Z}_0} \int_{\mathcal{Z}} s(z) \pi_Z(dz) \mu(dz_0) = \int_{\mathcal{Z}} s(z) \pi_Z(dz)$$

is finite by the V_Z–geometric ergodicity of Z_t. Because the integral $\int_{\mathcal{Z}_0} V_X(\lambda(z_0))\mu(dz_0)$ is also finite by assumption (b),

$$E_\mu[s(Z_n)] = \int_{\mathcal{Z}_0} \int_{\mathcal{Z}} s(z) P_Z^n(z_0, dz) \mu(dz_0)$$

is finite for all n and converges at geometric rate to the expectation taken with respect to the stationary distribution. This completes the proof. ■

Proof of Lemma 1. Consider the model (4)–(5) and suppose that the assumptions of Proposition 2 are satisfied apart from (c). As noticed in the discussion following (4)–(5), in this case the function λ is 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

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

For simplicity, denote $\int_{\mathcal{X}} \pi_X(dw) V_X(w) = C$ 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 $x \in \mathcal{X}$,

$$\begin{aligned} \left| \int_{\mathcal{X}} P_X(x, dw) V_X(w) \right| &\leq \left| \int_{\mathcal{X}} P_X(x, dw) V_X(w) - \int_{\mathcal{X}} \pi_X(dw) V_X(w) \right| + C \\ &\leq \varrho R V_X(x) + C \\ &\leq \max\{\varrho R, C\} (V_X(x) + 1) \\ &\leq 2 \max\{\varrho R, C\} V_X(x), \end{aligned}$$

where $\max\{\varrho R, C\} < \infty$. Combining the preceding inequalities therefore shows that assumption (c) holds. ■

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 of the theorem consists of proving 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)). These steps are standard in proving geometric ergodicity of a Markov chain. Irreducibility, the existence of a small set, and aperiodicity are first proven in Lemmas 3, 4 and 5, respectively.

Lemma 3 *Suppose that the assumptions of Theorem 1 hold. Then there exist real numbers* $\underline{l}$ *and* $\bar{l}$ *such that* $(\underline{l}, \bar{l}) \subset \mathcal{X}$ *and that the Markov chain* X_t *is* φ*–irreducible with* $\varphi(\cdot) = \mu_{Leb}(\cdot \cap (\underline{l}, \bar{l}))$.

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

$$E\left[(a + \epsilon + b(\varepsilon_t))^k\right] < 1. \tag{15}$$

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

$$f_{x2}(f_{y2}(x)\varepsilon_t, x) \leq xb(\varepsilon_t) + \frac{1}{2}\epsilon x \tag{16}$$

and

$$f_{x1}(x) \leq ax + \frac{1}{2}\epsilon x \tag{17}$$

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 (15) 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 5).

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

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

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

Establishing II and IV will complete the proof of Lemma 3, while III will be used later in the proof of Lemma 4.

Proof of I. Let $t \in \mathbb{Z}_+$ and suppose that $X_{t-1} \in S_{1\epsilon}$. Using (16) and (17) we find that

$$\begin{aligned} X_t &= f_{x1}(X_{t-1}) + f_{x2}(f_{y2}(X_{t-1})\varepsilon_{t-1}, X_{t-1}) \\ &\leq aX_{t-1} + \frac{1}{2}\epsilon X_{t-1} + X_{t-1}b(\varepsilon_{t-1}) + \frac{1}{2}\epsilon X_{t-1} \\ &= X_{t-1}\left(a + \epsilon + b(\varepsilon_{t-1})\right) \end{aligned}$$

and, since both sides are positive,

$$X_t^k \leq X_{t-1}^k \left(a + \epsilon + b(\varepsilon_{t-1})\right)^k.$$

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. Therefore, on the event Ω_n, we have

$$X_t^k \leq X_{t-1}^k \cdot E\left[(a+\epsilon+b(\varepsilon_{t-1}))^k\right], \tag{18}$$

where by (15) the expectation is < 1.

Now choose an arbitrary $x \in S_{1\epsilon}$, and denote $X_0 = x$. Using (18) inductively we have, for arbitrary $n \in \mathbb{Z}_+$ and on the event Ω_n, that

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

as long as $X_1, \ldots, X_{n-1} \in S_{1\epsilon}$. Since $E\left[(a+\epsilon+b(\varepsilon_{t-1}))^k\right] < 1$, the right-hand-side of (19) 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

$$\sup_{x \in S_{2\epsilon}} f_{x1}(x) \leq M_1 \quad \text{and} \quad \sup_{x \in S_{2\epsilon}} f_{x2}(R, x) \leq M_2. \tag{20}$$

Define $\underline{l} = \max\{M_1 + M_2, M_\epsilon + 1\}$, and choose an arbitrary $\bar{l} > \underline{l}$ (note that only the fact that $\underline{l} \geq M_1 + M_2$ is used in the present proof, while the fact that $\underline{l} > M_\epsilon$ is used later in the proof of Lemma 5).

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

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

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

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

From the assumed boundedness conditions for f_{x1}, f_{x2}, f_{y2}, $\partial f_{x2}^{-1}(v, x)/\partial v$, and ϕ_ε (bounded on bounded subsets for the first three functions, and suitably bounded away from zero for the last three functions) it follows that

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

for some positive ϵ_*, and therefore $P(x, A) \geq \epsilon_* \mu_{Leb}(A \cap (\underline{l}, \overline{l}))$. The set A can clearly be replaced by $A \cap (\underline{l}, \overline{l})$, and hence we find that

$$\inf_{x \in S_{2\epsilon}} P(x, A \cap (\underline{l}, \overline{l})) \geq \epsilon_* \mu_{Leb}(A \cap (\underline{l}, \overline{l})), \tag{21}$$

which establishes both II and III.

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 (21). This completes the proof of IV. ■

Lemma 4 *If the assumptions of Theorem 1 hold then the set $S_{2\epsilon}$ is small.*

Proof. Equation (21) shows that equation (5.14) of Meyn and Tweedie (1993) is satisfied 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 5 *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

$$\forall x \in A : \big(P(x, A) > 0 \text{ and } P^2(x, A) > 0\big) \tag{22}$$

for some small set A such that $\varphi(A) > 0$. In the following, we will prove that this holds with the set $(\underline{l}, \overline{l})$.

To this end, we first prove a result slightly stronger than needed, namely that for all open subsets A of $S_{1\epsilon}$ it holds that $P(x, A) > 0$ and $P^2(x, A) > 0$ for every $x \in A$. For this, choose an arbitrary open subset A of $S_{1\epsilon}$, and an arbitrary $x \in A$. Because $x > M_\epsilon$, we have by (17) and (15), that $f_{x1}(x) \leq \left(a + \frac{1}{2}\epsilon\right) x < x$, from which it also follows that $x - f_{x1}(x) \geq x - \left(a + \frac{1}{2}\epsilon\right) x = x(1 - a - \frac{1}{2}\epsilon)$. 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

$$\varepsilon \in (\underline{e}, \overline{e}) \Rightarrow f_{x1}(x) + f_{x2}(f_{y2}(x)\, \varepsilon, x) \in (x - \delta, x + \delta) \tag{23}$$

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

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

where the second inequality follows from (23) and the third from the assumed positivity of $\phi_\varepsilon(\cdot)$. In addition, by the Chapman-Kolmogorov equation,

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

Hence the assertion made is proven. Since this holds, in particular, for the set $(\underline{l}, \overline{l})$ the condition (22) is established with $A = (\underline{l}, \overline{l})$.

By Lemma 3, $\varphi\left((\underline{l}, \overline{l})\right) > 0$. To establish that the set $(\underline{l}, \overline{l})$ is small consider first the proof of parts II and III in Lemma 3 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

$$\inf_{x \in (\underline{l}, \overline{l})} P(x, A \cap (\underline{l'}, \overline{l'})) \geq \epsilon'_* \mu_{Leb}(A \cap (\underline{l'}, \overline{l'})) > 0 \tag{24}$$

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 by definition. This completes the proof of aperiodicity. ■

Finishing the proof of Theorem 1.

We have already proven that X_t is irreducible and aperiodic, and that the set $S_{2\epsilon}$ is small. Hence showing that condition (15.3) of Meyn and Tweedie (1993) holds for the function $V_X(x) = 1 + x^k$ shows that X_t is a V_X–geometrically ergodic Markov chain. The condition (15.3) holds if there exist constants $c_1 > 0$ and $c_2 < \infty$ such that

$$E\left[V_X(X_t) \mid X_{t-1} = x\right] \leq (1 - c_1) V_X(x) + c_2 \mathbf{1}\left(x \in S_{2\epsilon}\right) \tag{25}$$

for all $x \in \mathcal{X}$.

The expectation in (25) 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 3 we have

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

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 $E\left[(a+\epsilon+b(\varepsilon_{t-1}))^k\right] < 1$, it follows that

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

Defining $c_1 = \frac{1}{2}\left(1-E\left[(a+\epsilon+b(\varepsilon_{t-1}))^k\right]\right)$ shows that (25) holds for all $x \in S_{1\epsilon}$.

Suppose now that $x \in S_{2\epsilon}$. By the first inequality in (20) and Assumption 2(e) we have

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

Defining $c_2 = 1+E\left[(M_1+c+M_\epsilon b(\varepsilon_{t-1}))^k\right]$ and noting that $(1-c_1)V(x)$ is always positive shows that (25) 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 $\int_{\mathcal{Y}} \pi_{Y|X}(dy|x)V_Z(y,x) \le cV_X(x)$ 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

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

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

$$
\begin{aligned}
\int_{\mathcal{Y}} \pi_{Y|X}(dy|x)V_Z(y,x) &= 1+x^k+E|x^{1/d}\varepsilon_t+f_{y1}(x)|^{dk} \\
&\le 1+x^k+\left(\left(E|x^{1/d}\varepsilon_t|^{dk}\right)^{1/dk}+\left(|f_{y1}(x)|^{dk}\right)^{1/dk}\right)^{dk} \\
&\le 1+x^k+C_1\left(\left(E|x^{1/d}\varepsilon_t|^{dk}\right)+|f_{y1}(x)|^{dk}\right) \\
&\le 1+x^k+C_1\left(x^kE|\varepsilon_t|^{dk}+(\mu_0+\mu_1x^{1/e})^{dk}\right) \\
&\le 1+x^k+C_1\left(x^kE|\varepsilon_t|^{dk}+C_2(\mu_0^{dk}+\mu_1^{dk}x^{dk/e})\right) \qquad (26)
\end{aligned}
$$

for some constants C_1 and C_2. Here the first inequality follows from the Minkowski's inequality, and the second and fourth from the fact that for any $p \ge 1$ there exists a constant C such that $(a+b)^p \le C(a^p+b^p)$ for all non-negative numbers a and b (this follows from the equivalence of the l_1– and l_p–norms in finite dimensional vector spaces when $p \geq 1$, and is applied above with $p = dk$). In (26), $x^{dk/e} \leq \max\{1, x^k\} \leq (1 + x^k)$, and hence the expression in (26) is smaller than $C_3(1 + x^k)$ for some constant C_3.

The case $d > e$ can be proven similarly and it suffices to note that

$$\begin{aligned} E|x^{1/d}\varepsilon_t + f_{y1}(x)|^{ek} &\leq \left(\left(E|x^{1/d}\varepsilon_t|^{ek}\right)^{1/ek} + \left(|f_{y1}(x)|^{ek}\right)^{1/ek}\right)^{ek} \\ &\leq C_1\left(x^{ek/d}E|\varepsilon_t|^{ek} + (\mu_0 + \mu_1 x^{1/e})^{ek}\right) \\ &\leq C_1\left(x^{ek/d}E|\varepsilon_t|^{ek} + C_2(\mu_0^{ek} + \mu_1^{ek}x^k)\right) \\ &\leq C_1\left((1 + x^k)E|\varepsilon_t|^{ek} + C_2(\mu_0^{ek} + \mu_1^{ek}x^k)\right) \\ &\leq C_3(1 + x^k). \end{aligned}$$

Hence, the proof is complete. ■

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 (25)). 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 the TACD–model. Denoting $R_j = [r_{j-1}, r_j)$ the TACD model can be written as

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

Defining $f_{x1}(x) = \beta_J x$ and

$$f_{x2}(x\varepsilon, x) = \sum_{j=1}^{J} (\omega_j + \alpha_j x\varepsilon + \beta_j x) \mathbf{1}(x\varepsilon \in R_j) - \beta_J x$$

we have $X_t = f_{x1}(X_{t-1}) + f_{x2}(X_{t-1}\varepsilon_{t-1}, X_{t-1})$. 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 $f_{x2}([r_{J-1}, \infty), x) = [\omega_J + \alpha_J r_{J-1}, \infty)$ for all x.

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

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

Therefore

$$
\begin{aligned}
f_{x2}(x\varepsilon, x) &= \sum_{j=1}^{J} (\omega_j + \alpha_j x\varepsilon + \beta_j x) 1(x\varepsilon \in R_j) - \beta_J x \\
&\leq \max \omega_j + r_{J-1} \max \alpha_j + M \max \beta_j + x \max \beta_j \mathbf{1}(\varepsilon < r_{J-1}/M) + \alpha_J x\varepsilon \\
&= (\max \omega_j + r_{J-1} \max \alpha_j + M \max \beta_j) + x(\alpha_J \varepsilon + \max \beta_j \mathbf{1}(\varepsilon < r_{J-1}/M))
\end{aligned}
$$

and, denoting $c = (\max \omega_j + r_{J-1} \max \alpha_j + M \max \beta_j)$ and $b(\varepsilon) = (\alpha_J \varepsilon + \max \beta_j \mathbf{1}(\varepsilon < r_{J-1}/M))$, 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 $k \geq 1$, if $E(\alpha_J \varepsilon_t + \beta_J)^k < 1$, 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 $k = 1$, $E(a + b(\varepsilon_t)) = E(\alpha_J \varepsilon_t + \beta_J) + \max \beta_j E\mathbf{1}(\varepsilon_t < r_{J-1}/M)$. By choosing M sufficiently large, the last term can be made arbitrarily small, and hence $E(a + b(\varepsilon_t)) < 1$ for a suitable choice of M. Suppose now that $k > 1$. By Minkowski's inequality we have

$$
\begin{aligned}
E(a + b(\varepsilon_t))^k &= E((\alpha_J \varepsilon_t + \beta_J) + \max \beta_j \mathbf{1}(\varepsilon < r_{J-1}/M))^k \\
&\leq \left(\left(E(\alpha_J \varepsilon_t + \beta_J)^k \right)^{1/k} + \left(E(\max \beta_j \mathbf{1}(\varepsilon < r_{J-1}/M))^k \right)^{1/k} \right)^k .
\end{aligned}
$$

Here the second expectation can be written as $E(\max \beta_j \mathbf{1}(\varepsilon < r_{J-1}/M))^k = \max \beta_j^k \cdot E[\mathbf{1}(\varepsilon_t < r_{J-1}/M)]$ and hence, by choosing M sufficiently large, this term can be made small enough so that $E(a + b(\varepsilon_t))^k < 1$. This completes establishing the validity of the conditions. ■

References

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.

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.

COMTE, F., AND O. LIEBERMAN (2003): "Asymptotic theory for multivariate GARCH processes," *Journal of Multivariate Analysis*, 84, 61–84.

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

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

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 (in press): "A family of autoregressive conditional duration models," *Journal of Econometrics*, Working paper version available from *http://epge.fgv.br/portal/arquivo/1392.pdf.*

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.

LANNE, M., AND P. SAIKKONEN (2004): "Nonlinear GARCH models for highly persistent volatility," Manuscript, University of Helsinki.

LIEBSCHER, E. (2004): "Towards a unified approach for proving geometric ergodicity and mixing properties of nonlinear autoregressive processes," Manuscript, Technical University Ilmenau.

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.

Meitz, M., and T. Teräsvirta (2004): "Evaluating models of autoregressive conditional duration," SSE/EFI Working Paper Series in Economics and Finance No. 557, available from *http://swopec.hhs.se/hastef/abs/hastef0557.htm*.

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.

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.