Electronic Journal of Statistics
Vol. 8 (2014) 1088–1112
ISSN: 1935-7524
DOI: 10.1214/14-EJS924

Stationarity and ergodicity of univariate generalized autoregressive score processes*

Francisco Blasques[†], Siem Jan Koopman[‡] and André Lucas[†]

VU University Amsterdam
de Boelelaan 1105, 1081 HV Amsterdam, The Netherlands
e-mail: f.blasques@vu.nl; s.j.koopman@vu.nl; a.lucas@vu.nl

Abstract: We characterize the dynamic properties of generalized autoregressive score models by identifying the regions of the parameter space that imply stationarity and ergodicity of the corresponding nonlinear time series process. We show how these regions are affected by the choice of parameterization and scaling, which are key features for the class of generalized autoregressive score models compared to other observation driven models. All results are illustrated for the case of time-varying means, variances, or higher-order moments.

MSC 2010 subject classifications: Primary 60G10, 62M10; secondary 91B84.
Keywords and phrases: Nonlinear dynamics, observation driven time-varying parameter models, stochastic recurrence equations, contracting properties.

Received January 2014.

Contents

*We would like to thank Peter Boswijk, Richard Davis, Sara van de Geer, Andrew Harvey, Anders Rahbek, seminar participants at Tinbergen Institute seminars, and participants of the Amsterdam-Cambridge Conference on Score Driven Models, January 2013, for useful discussions.

[†]Blasques and Lucas thank the Dutch National Science Foundation (NWO; grant VICI453-09-005) for financial support.

[‡]Koopman acknowledges support from CREATES, Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation.

1. Introduction

Time-varying parameter models are often used in empirical econometric analysis. Cox [3] established the nomenclature for the two alternative frameworks for capturing time-varying parameters, namely observation driven and parameter driven time series models. In this paper we concentrate on the class of observation driven models, which includes famous examples such as the autoregressive conditional heteroskedasticity (ARCH) model of Engle [9]. The main advantage of observation driven models is that the likelihood is available in closed form, such that in many cases parameter estimation is relatively simple to implement. A unified observation driven framework for time-varying parameters applicable in a general parametric context, however, was lacking thus far.

In a sequence of recent papers, Creal et al. (2011, 2013a,b) [4, 5, 6] introduced a new unified class of observation driven time-varying parameter models, called Generalized Autoregressive Score (GAS) models. GAS models combine in a consistent manner the dynamics of time-varying parameters and the conditional distribution of the observed data by driving the time-varying parameter using the scaled score of the conditional density. Creal et al. show that GAS models encompass many well-known observation driven time series models, including the ARCH model of Engle [9], the generalized ARCH (GARCH) model of Bollerslev [1], the exponential GARCH (EGARCH) model of Nelson [24], the autoregressive conditional duration (ACD) model of Engle and Russell [11], the multiplicative error model (MEM) of Engle [10], the autoregressive conditional multinomial (ACM) model of Rydberg and Shephard [25], the Beta-t-GARCH model of Harvey [15], and many related models. In addition, the GAS framework gives rise to many new time-varying parameter models.

A simple illustration of the usefulness of GAS models is given in Figure 1 for a time-varying volatility model for fat-tailed financial return data. The figure displays the absolute returns in Norwegian Krone on Nordpool electricity prices, together with two estimated volatility series based on GAS models with a normal and a Student's t distribution, respectively. The dashed curve in Figure 1 gives the estimated volatility series based on a GAS model and a conditional normality assumption. The GAS model in this case coincides with the standard

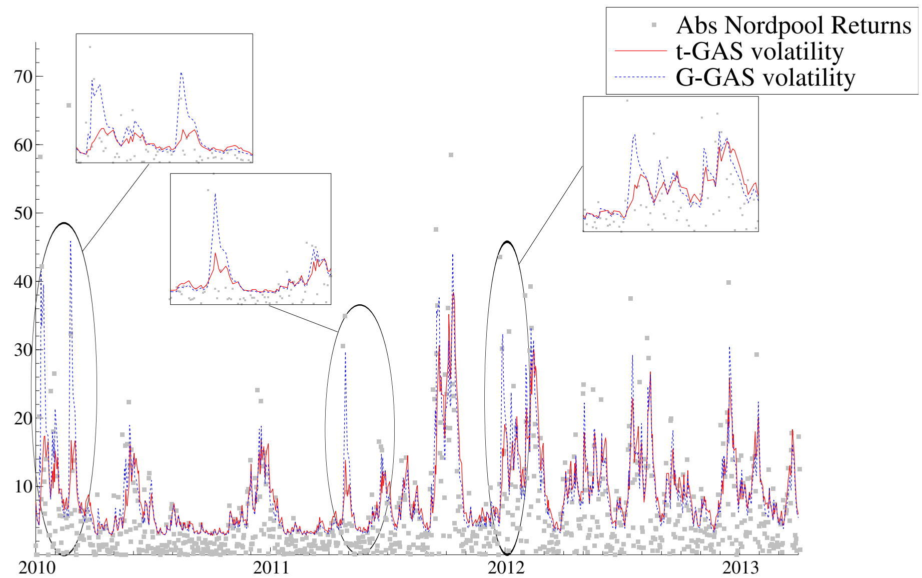

Fig 1. *Time series of returns on Nordpool electricity prices and volatility estimates from the GAS model under the assumption of a Gaussian distribution (dashed curve) and a Student's t distribution (solid curve). Data obtained from Datastream.*

GARCH(1,1) model of Bollerslev [1]. As the normal distribution does not have fat tails, the model dynamics imply that a large absolute return must be fully attributed to an increase in volatility. This causes the spikes in the figure, followed by the rapid subsequent decay, see for example the zoomed-in plots for January and February 2010, April 2011, and December 2011. The solid curve in the figure is obtained for a GAS model under the assumption of a conditional Student's t distribution. This model is very different from the familiar GARCH-t model of Bollerslev [1], and is in fact much closer to the recent Beta-t-GARCH model of Harvey [15]. In particular, the model uses the score of the Student's t distribution rather than the lagged squared returns to drive the volatility dynamics; see also Section 2. The different effect on how the volatility estimates respond to large absolute returns is apparent, particularly in the zoomed-in panels in Figure 1. The volatility estimates from the GAS model based on the Student's t distribution recognize that large absolute returns need not (only) be due to volatility increases, but can also result from tail realizations of the Student's t distribution. As a result, we obtain much less erratic volatility dynamics. More examples can be found in Creal et al. (2013a,b) [5, 6], Harvey [15], Lucas et al. [22], and Harvey and Luati [17].

Despite the proven empirical usefulness of GAS models and their conceptual generalizability, relatively little is known about the general stochastic behavior of GAS processes except for some well-known special cases such as the ARCH and GARCH models and the Beta-t-GARCH model of Harvey [15]. In this paper we address this issue by providing explicit conditions for stationarity and ergodicity for a large class of GAS processes. In particular, we give a characterization of the region of the parameter space that renders the process stationary and ergodic. Establishing stationarity and ergodicity not only allows for a better understanding of the nature of the GAS process, it also plays an important role in standard proofs of estimator consistency and asymptotic normality. Indeed, the stationarity and ergodicity of the underlying time series process, together with appropriate model invertibility, bounded moments and smoothness conditions, allows for the properties of the maximum likelihood estimator to be obtained by appealing to appropriate laws of large numbers and central limit theorems. The stationarity and ergodicity properties are the focal point of the current paper, while establishing the existence of moments and invertibility to future work. For a careful discussion about the importance of the latter, we refer to Wintenberger [28] for the exponential GARCH case.

Our approach builds on the contraction condition formulated in Bougerol [2] and Straumann and Mikosch [27] for general stochastic recurrence equations. As shown in Straumann and Mikosch [27], this contraction condition can be used to ensure model invertibility when applied to the filtering equation. Here we focus on using the contraction condition to ensure the strict stationarity and ergodicity (SE) of the GAS model as a data generating process. In particular, we derive sufficient conditions for the supremum Lipschitz constants of the GAS (stochastic) recurrence relations to be bounded in expectation. The nonlinearity of the GAS recursions only allows us to establish sufficient conditions for strict stationarity and ergodicity. If the GAS recursion collapses to a linear form, stronger results might be obtainable since the contraction condition becomes both necessary and sufficient for SE; see Nelson [23] for results related to the GARCH model and Jensen and Rahbek [19] who explore such results to derive ML asymptotics for the nonstationary GARCH. Even in such cases, however, not all parameters in the model may be estimated consistently. This is true in particular for the intercept in the GARCH update equation; see Francq and Zakoian [12].

A complication is provided by the generality of the GAS framework which allows one to select the distribution of the data, the parameterization of the time-varying parameter, and the scaling of the score function that governs the dynamic processes of the parameters. Each of these choices yields a different model and is directly relevant for the SE properties of the dynamic parameter. In particular, these choices determine also the region of the parameter space that renders the process strictly stationary and ergodic. We call this the SE region of the parameter space.

The remainder of this paper is organized as follows. In Section 2 we introduce the GAS model, its parameterization, and its scaling. In Section 3 we derive our characterization of stationarity and ergodicity regions and provide some generic examples. In Section 4 we provide a range of concrete GAS models for time-varying means, variances, and higher-order moments to show how the results from Section 3 can actually be applied. We conclude in Section 5. The Appendix gathers the proofs.

2. The generalized autoregressive score model

Consider a real-valued stochastic sequence of observations $\{y_t\}_{t\in\mathbb{Z}}$ with conditional probability density,

$$p_y\Big(y_t \;\Big|\; h\big(f_t(\theta)\big);\ \lambda\Big), \tag{2.1}$$

for all $t \in \mathbb{Z}$, where $\{f_t(\theta)\}_{t\in\mathbb{Z}}$ represents a scalar time-varying parameter that depends on a vector of time-invariant parameters $\theta \in \Theta$, $h : \mathbb{R} \to \mathbb{R}$ is a link function, and $\lambda \in \Lambda$ is a vector of time-invariant parameters that indexes the conditional density p_y. Equation (2.1) contains many models of empirical interest. For example, λ may denote the degrees of freedom parameter of a Student's t distribution, h the identity function, and $\{f_t(\theta)\}_{t\in\mathbb{Z}}$ a time-varying variance to obtain a time-varying volatility model for the Student's t distribution. The model in (2.1) can be further extended to allow for exogenous variables, lagged endogenous variables, and lagged values of $f_t(\theta)$ in the conditioning set; see Creal et al. (2011, 2013a) [4, 5].

The Generalized Autoregressive Score (GAS) framework specifies the dynamic process for the time-varying parameter $\{f_t(\theta)\}_{t\in\mathbb{Z}}$ as

$$\begin{aligned} f_{t+1}(\theta) &= \omega + \alpha s_t(f_t(\theta);\lambda) + \beta f_t(\theta), && (2.2)\\ s_t\big(f_t(\theta);\lambda\big) &= S\big(f_t(\theta);\lambda\big)\cdot\nabla_t\big(f_t(\theta);\lambda\big), && (2.3)\\ \nabla_t\big(f_t(\theta);\lambda\big) &= \partial\log p_y\big(y_t|f;\lambda\big)\big/\,\partial f\big|_{f=f_t(\theta)}, && (2.4) \end{aligned}$$

where ω, α, and β are time-invariant parameters, $S(f_t(\theta);\lambda)$ is a univariate scaling factor for the score $\nabla_t(f_t(\theta);\lambda)$ of the conditional observation density (2.1), and log denotes the natural logarithm. The current GAS model specification has one lag of $s_t(f_t(\theta);\lambda)$ and one lag of $f_t(\theta)$ on the right-hand side of (2.2). The inclusion of more lags for $f_t(\theta)$ or $s_t(f_t(\theta);\lambda)$ is straightforward; see Creal et al. (2013a) [5] for more details. We define the parameter vector $\theta \in \Theta$ as $\theta = (\omega, \alpha, \beta, \lambda')'$, with Θ denoting the parameter space.

The key element in (2.2) is the definition of $s_t(f_t(\theta);\lambda)$ as the scaled score of the conditional observation density in (2.1) with respect to the time-varying parameter $f_t(\theta)$. The intuition for this is straightforward: at time t we improve the local fit of the model as measured by the log conditional observation density $\log p_y(y_t|f_t(\theta);\lambda)$. We do so by taking a scaled step in the steepest ascent direction of the model's fit at time t. Since $s_t(f_t(\theta);\lambda)$ is a function of past data and parameters alone, the GAS model can be classified as an observation driven model; see Cox [3].

An advantage of the choice of $s_t(f_t(\theta);\lambda)$ as the driving mechanism in (2.2) is that it can be applied whenever an explicit expression for the conditional observation density is available. The equations (2.1) to (2.4) therefore encompass a large set of familiar time series models. For example, if p_y is the normal density, $f_t(\theta)$ is the time-varying variance, and h is the identity function, we obtain the standard GARCH model of Bollerslev [1]; see also the discussion and references in Section 1. For other choices, the GAS framework gives rise to entirely new time-varying parameter models, the dynamic properties of which have typically not been studied before; see Creal et al. (2013b) [6] for an elaborate example in the credit risk setting.

Each choice for the scaling function $S = S(f_t(\theta); \lambda)$ in (2.3) gives rise to a new GAS model. Intuitive choices for S may relate to the local curvature of the score as measured by the inverse information matrix, for example

$$S(f_t(\theta); \lambda) = (\mathcal{I}_t(f_t(\theta); \lambda))^{-a}, \tag{2.5}$$

where

$$\mathcal{I}_t(f_t(\theta); \lambda) = \mathrm{E}_{t-1}[\nabla_t(f_t(\theta); \lambda)\nabla_t(f_t(\theta); \lambda)'],$$

and where a is typically taken as 0, 1/2, or 1. Other choices of S are possible as well.

Similarly, the choice of a link function h provides another degree of freedom for model specification. For example, in a time-varying variance setting, we can choose to model the variance directly by setting $h(f_t(\theta)) = f_t(\theta)$ with $f_t(\theta)$ representing the variance. Alternatively, we can opt for modelling the log variance by setting $h(f_t(\theta)) = \exp(f_t(\theta))$ with $f_t(\theta)$ representing the log variance. The latter specification can have the advantage that the variance itself is always positive by construction, even if $f_t(\theta)$ becomes negative.

To provide further structure to the probability density function in (2.1) we let it be implicitly defined by the following observation equation,

$$y_t = g_\lambda\big(h(f_t(\theta)), u_t\big) \; \forall\, t \in \mathbb{Z}, \tag{2.6}$$

where for all $\lambda \in \Lambda$, $g_\lambda : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$ is a function and $\{u_t\}_{t\in\mathbb{Z}}$ is an independently identically distributed sequence with u_t independent of f_t for every t and $u_t \sim p_{u,\lambda}(u_t)$. This structure covers many cases of empirical interest. For example, a time-varying volatility model is obtained by setting $y_t = f_t(\theta)^{1/2} \cdot u_t$, with $f_t(\theta)$ denoting the time-varying variance and u_t being, for example, normally or Student's t distributed. Many other models are contained in (2.6) as well by letting g_λ be the inverse distribution function corresponding to (2.1), and by letting u_t be a uniform random variable on $[0, 1]$. For example, a model for a Student's t distribution with time-varying degrees of freedom parameter is captured by taking u_t as a uniform, h as the identity function, and g_λ as the inverse Student's t distribution function with $f_t(\theta)$ degrees of freedom.

The stochastic properties of $\{y_t\}_{t\in\mathbb{Z}}$ are now fully determined by *(i)* the parameterization h, *(ii)* the family of densities $p_u = \{p_{u,\lambda}\}_{\lambda\in\Lambda}$, *(iii)* the family of transformation functions $g = \{g_\lambda\}_{\lambda\in\Lambda}$, *(iv)* the scaling function S, and *(v)* the parameter value $\theta \in \Theta$. In other words, a probability measure for $\{y_t\}$ is defined whenever a point

$$(h, p_u, g, S, \theta) \;\in\; \mathcal{H} \times \mathcal{P}_u \times \mathcal{G} \times \mathcal{S} \times \Theta$$

is selected, where $\mathcal{H}$ denotes the space of link functions, $\mathcal{P}_u$ the space of families of densities p_u for u_t, $\mathcal{G}$ the space of families of transformation functions g, and $\mathcal{S}$ the space of scaling functions. Given this notation, we can now start characterizing stationarity and ergodicity regions for GAS processes.

3. Stationarity and ergodicity

3.1. Stochastic recurrence equations

To characterize the dynamic properties of GAS processes, we use the stationarity and ergodicity conditions formulated by Bougerol [2] and Straumann and Mikosch [27] for general stochastic recurrence equations; see also Diaconis and Freedman [7] and Wu and Shao [29]. In particular, we define subsets of $\mathcal{H} \times \mathcal{P}_u \times \mathcal{G} \times \mathcal{S} \times \Theta$ that render $\{y_t\}_{t\in\mathbb{Z}}$ stationary and ergodic (SE). Measurability of the relevant maps is implied by explicit assumptions about continuity of the relevant maps and by letting the relevant domain and image spaces be measurable sets equipped with Borel σ-algebras generated by the topology of each respective set.

Let $f_t(f_1,\theta)$ denote the value at time $t \in \mathbb{N}$ of the time-varying parameter, with random initialization f_1 taking values in $\mathcal{F} \subseteq \mathbb{R}$ and dynamics determined by $\theta \in \Theta$. A stochastic recurrence equation for the sequence $\{f_t(f_1,\theta)\}_{t\in\mathbb{N}}$ takes the form

$$f_1(f_1,\theta) = f_1 \quad \text{and} \quad f_{t+1}(f_1,\theta) = \phi_t\big(f_t(f_1,\theta);\theta\big) \quad \forall\, t \in \mathbb{N}, \tag{3.1}$$

where $\phi_t : \mathcal{F} \to \mathcal{F}$ is a random function. This clearly embeds the GAS model in (2.2) with random initialization f_1 in $\mathcal{F}$ by setting

$$\phi_t\big(f_t(f_1,\theta);\theta\big) = \omega + \alpha s_t\big(f_t(f_1,\theta);\lambda\big) + \beta f_t(f_1,\theta)\ \forall\, t \in \mathbb{N}, \tag{3.2}$$

with every y_t in $s_t\big(f_t(f_1,\theta);\lambda\big)$ replaced by $g_\lambda\big(h(f_t(f_1,\theta)),u_t\big)$ from equation (2.6). Sufficient conditions for $\{f_t(f_1,\theta)\}_{t\in\mathbb{N}}$ to converge exponentially almost surely (e.a.s.) to a unique SE sequence $\{f_t(\theta)\}_{t\in\mathbb{Z}}\ \forall\, \theta \in \Theta$ are given below.[1]

Assumption 1. *For every $\theta \in \Theta$, $\{\phi_t(\cdot;\theta)\}_{t\in\mathbb{Z}}$ is an SE sequence of Lipschitz maps $\phi_t(\cdot;\theta) : \mathbb{R} \to \mathbb{R}$ satisfying $\mathrm{E}\left[\log^+|\phi_0(f;\theta) - f|\right] < \infty$ for some $f \in \mathcal{F}$, with $\log^+(x) = \max(0,\log(x))$.*

Assumption 2. *For every $\theta \in \Theta$, the sequence $\{\phi_t(\cdot;\theta)\}_{t\in\mathbb{Z}}$ satisfies*

$$\mathrm{E}\log\left[\sup_{(f,f')\in\mathcal{F}\times\mathcal{F}:\ f\neq f'} \frac{|\phi_0(f;\theta) - \phi_0(f';\theta)|}{|f-f'|}\right] < \infty, \tag{3.3}$$

and

$$\mathrm{E}\log\left[\sup_{(f,f')\in\mathcal{F}\times\mathcal{F}:\ f\neq f'} \frac{|\phi_0^{(r)}(f;\theta) - \phi_0^{(r)}(f';\theta)|}{|f-f'|}\right] < 0, \tag{3.4}$$

for some $r \geq 1$ with $\phi_0^{(r)} = \phi_0 \circ \cdots \circ \phi_{-r+1}$.

The proof of the following lemma can be found in Bougerol [2]. Uniqueness of the limit process is added in Straumann and Mikosch [27].

[1] A sequence $\{x_t\}_{t\in\mathbb{N}}$ is said to converge e.a.s. if there exists $\gamma > 1$ such that $\gamma^t|x_t| \xrightarrow{p} 0$.

Lemma 1. *Let Assumptions 1 and 2 hold for a real-valued sequence* $\{f_t(f_1,\theta)\}_{t\in\mathbb{N}}$ *generated by* (3.1). *Then for every* $\theta \in \Theta$ *there exists a unique real-valued SE sequence* $\{f_t(\theta)\}_{t\in\mathbb{Z}}$ *such that* $|f_t(f_1,\theta) - f_t(\theta)| \overset{e.a.s.}{\to} 0$ *as* $t \to \infty$.

The SE properties for $\{y_t\}_{t\in\mathbb{Z}}$ defined in (2.6) follow directly from those of $\{f_t(\theta)\}_{t\in\mathbb{Z}}$. This is stated in the following assumption and proposition.

Assumption 3. *The sequence* $\{u_t\}_{t\in\mathbb{Z}}$ *in* (2.6) *consists of independently identically distributed (i.i.d.) random variables and the function* g_λ *is continuous for every* $\lambda \in \Lambda$.

Proposition 1. *Let Assumptions 1–3 hold and let the real-valued sequence* $\{f_t(\theta)\}_{t\in\mathbb{Z}}$ *be the unique SE solution to* $\{f_t(f_1,\theta)\}_{t\in\mathbb{N}}$ *generated by (3.1). Let the link function* $h : \mathbb{R} \to \mathbb{R}$ *be continuous. Then* $\{y_t\}_{t\in\mathbb{Z}}$ *in (2.6) is an SE random sequence for every* $\theta \in \Theta$.

The proposition can also be obtained under much weaker conditions on the sequence $\{u_t\}_{t\in\mathbb{Z}}$, such as stationarity and ergodicity, but for our current expositional purposes the assumption of an independently identically distributed $\{u_t\}_{t\in\mathbb{Z}}$ suffices.

In practice, condition (3.4) in Assumption 2 is the most challenging, both in terms of the class of models that it restricts, and in terms of analytic verification. It might be tempting to use numerical methods to evaluate condition (3.4) in applications. However, this approach is complicated by the fact that taking the expectation of the supremum of a possibly highly nonlinear function can lead to misleading results. Mistakes can then easily be made if the numerical algorithm for computing the supremum fails to find a global maximum for every $\theta \in \Theta$ and every $u_t \in \mathbb{R}$. To prevent such mistakes, we shall provide an analytical characterization of these SE regions. The following immediate result is helpful in this respect.

Proposition 2. *Let* ϕ_t *be given by (3.2) and* $s_t(f;\lambda)$ *be almost surely (a.s.) continuously differentiable in* f. *Then under Assumption 3, conditions (3.3) and (3.4) in Assumption 2 are implied by*

$$\mathrm{E} \sup_{f^*\in\mathcal{F}} \left| \beta + \alpha \frac{\partial s_t(f^*;\lambda)}{\partial f} \right| < 1, \tag{3.5}$$

which in turn is implied by

$$\mathrm{E} \sup_{f^*\in\mathcal{F}} \left| \frac{\partial s_t(f^*;\lambda)}{\partial f} \right| < \frac{1-|\beta|}{|\alpha|}. \tag{3.6}$$

In applications, both conditions (3.5) and (3.6) play an important role in the characterization of the SE region. We illustrate this in Section 4.

From condition (3.6) it follows directly that the sufficient SE region:

(i) consists at most of the interval $(-1, 1)$ in the direction of β; and
(ii) consists of an interval (α^-, α^+) with $\alpha^- \le 0 \le \alpha^+$ in the direction of α.

The supremum in (3.6) also reveals that if $\mathcal{F}$ depends on ω, the value of ω can influence the SE region. We shall deal conservatively with such cases by adopting always the largest possible $\mathcal{F}$. As a result, we focus our discussion entirely on the size of the SE region in terms of (α, β, λ). In particular, for any given value of λ, we obtain an SE region in the (α, β)-plane. Clearly, the maximum SE region dictated by Proposition 2 is obtained if

$$\mathrm{E} \sup_{f^* \in \mathcal{F}} |\partial s_t(f^*; \lambda)/\partial f| = 0.$$

In this case, only condition (i) above is binding, while $\alpha^- = -\infty$, and $\alpha^+ = \infty$. In other words, when the score is zero, then α can take any value as it does not affect the contraction condition, and hence condition (ii) becomes irrelevant. The SE region for given λ then becomes a rectangle of infinite length in the (α, β)-plane as characterized by $|\beta| < 1$, irrespective of the value of α and ω. If, on the other hand,

$$\mathrm{E} \sup_{f^* \in \mathcal{F}} |\partial s_t(f^*; \lambda)/\partial f| = \infty,$$

we obtain the degenerate SE region in the (α, β)-plane which we can characterize by $\{(\alpha, \beta) \mid |\beta| < 1, \alpha = 0\}$. The intermediate cases are characterized by the condition

$$0 < \mathrm{E} \sup_{f^* \in \mathcal{F}} |\partial s_t(f^*; \lambda)/\partial f| < \infty.$$

In such an intermediate case, we obtain a non-degenerate, bounded SE region in (α, β) for every given $\lambda \in \Lambda$. Given the structure of equation (3.6), such a region takes the form of a composition of triangles.

In all cases the actual SE region can be larger due to the fact that (3.6) only provides sufficient conditions for SE. We come back to this in the concrete examples in Section 4. The current set of conditions, however, already provides considerable insight into the type of SE regions that can be obtained for various types of GAS models. The conditions simplify for the special case where $s_t(\cdot; \lambda)$ is a linear function of $f_t(f_1, \theta)$. This includes a number of GAS models for time-varying volatilities and means. These GAS models behave substantially different from their GARCH counterparts; see Creal et al. [4]. In particular, their SE properties and conditions have as yet not been fully investigated. Only the case of the volatility model with $h(\cdot) = \log(\cdot)$, $S(\cdot; \lambda) = 1$, and p_y the Student's t density has been investigated by Harvey [15]. The results for an $s_t(\cdot; \lambda)$ that is affine in $f_t(f_1, \theta)$ are summarized in the following corollary.

Corollary 1. *Let* $s_t(f_t(f_1, \theta); \lambda) = \zeta_{1,t}(\lambda) \cdot f_t(f_1, \theta) + \zeta_{2,t}(\lambda)$, *where* $\zeta_{1,t}(\lambda)$ *and* $\zeta_{2,t}(\lambda)$ *are real-valued random variables. A sufficient condition for SE is then given by* $\mathrm{E}|\zeta_{1,t}(\lambda)| < (1 - |\beta|)/|\alpha|$.

The corollary makes clear that if s_t is affine in $f_t(f_1, \theta)$, the maximal SE region is obtained if $\zeta_{1,t}(\lambda) \equiv 0$. A non-degenerate SE region is obtained whenever $\mathrm{E}|\zeta_{1,t}(\lambda)| < \infty$. A smaller value for this expectation ensures a larger SE region in the (α, β)-plane. If $\mathrm{E}|\zeta_{1,t}(\lambda)|$ is unbounded, we obtain the degenerate SE region.

3.2. Non-degeneracy and bounds for SE regions

We separate the analysis of SE regions two steps. First, we determine whether there exists some λ for which the SE region is non-degenerate in the (α,β)-plane. Second, we determine the SE region's actual size and shape in terms of $(\omega,\alpha,\beta,\lambda)$. While the first step is more limited, it already allows for a characterization of the stochastic properties of GAS processes over a range of parameters (ω,α,β) determining the dynamics of $\{f_t(\theta)\}$ and $\{y_t\}$ for a given conditional density. The second step extends the analysis by characterizing stochastic properties over a family of conditional densities indexed by λ.

As shown in Section 3.1, the existence and size of a non-degenerate SE region both depend on the value of $\mathrm{E}\sup_{f^*\in\mathcal{F}}|\partial s_t(f^*;\lambda)/\partial f|$. In what follows, we obtain meaningful upper bounds for this expectation for any combination of a given link function h, family of distributions p_u, and scale function S that generates a separable upper bound.

Assumption 4. *Let s_t be a.s. continuously differentiable with*

$$|\partial s_t(f;\lambda)/\partial f| \quad\leq\quad |\eta(f;\lambda)\zeta_1(u_t;\lambda)+\zeta_2(u_t;\lambda)| \ \ \forall\ f\in\mathcal{F}, \tag{3.7}$$

where $\sup_{f\in\mathcal{F}}|\eta(f;\lambda)|\leq\bar{\eta}(\lambda)<\infty$, $\mathrm{E}|\zeta_1(u_t;\lambda)|\leq\bar{\zeta}_1(\lambda)<\infty$, *and* $\mathrm{E}|\zeta_2(u_t;\lambda)|\leq\bar{\zeta}_2(\lambda)<\infty$.

Proposition 3 below gives the main result. When Λ is a singleton in $\mathbb{R}^q$, we mainly use the proposition for models where the distribution of u_t is assumed to be known (λ is given). When λ is unknown, then we use the proposition with Λ being a non-degenerate subset of $\mathbb{R}$. In these models we assume the distribution of u_t belongs to some parametric family of distributions indexed by $\lambda\in\Lambda$ and we obtain the properties of the stochastic sequence with λ ranging over an appropriately defined set Λ.

Proposition 3. *Let h be continuous and $\{u_t\}_{t\in\mathbb{Z}}$ in (2.6) be an i.i.d. sequence. For every $\lambda\in\Lambda$ suppose that $\mathrm{E}\left[\log^+|s_0(f;\lambda)-f|\right]<\infty$ for some $f\in\mathcal{F}$, g_λ is continuous, and Assumption 4 holds. Then $\mathrm{E}\sup_{f^*}|\partial s_t(f^*;\lambda)/\partial f|<\infty$ and $\{f_t(f_1,\theta)\}_{t\in\mathbb{N}}$ generated by (3.1) converges e.a.s. to the unique SE solution $\{f_t(\theta)\}_{t\in\mathbb{Z}}$ and $\{y_t\}_{t\in\mathbb{Z}}$ defined in (2.6) is an SE random sequence for every $(\omega,\alpha,\beta,\lambda)\in\Theta$ where*

$$\Theta := \left\{(\omega,\alpha,\beta,\lambda)\in\mathbb{R}\times\mathbb{R}\times(-1,1)\times\Lambda\ :\ |\alpha|<\frac{1-|\beta|}{\bar{\eta}(\lambda)\times\bar{\zeta}_1(\lambda)+\bar{\zeta}_2(\lambda)}\right\}.$$

As we shall see in the examples in Section 4, Proposition 3 allows us to identify SE regions of the parameter space in a wide class of GAS models with nonlinear dynamics. We also note the following. For $s_t(f_t(f_1,\theta);\lambda) = S(f_t(f_1,\theta);\lambda)\cdot\nabla_t(f_t(f_1,\theta);\lambda)$, we can rewrite (3.5) and (3.6) of Proposition 2 as

$$\mathrm{E}\log\sup_{f^*\in\mathcal{F}}\left|\beta+\alpha\left(S(f^*;\lambda)\frac{\partial\nabla_t(f^*;\lambda)}{\partial f}+\frac{\partial S(f^*;\lambda)}{\partial f}\cdot\nabla_t(f^*;\lambda)\right)\right|<0, \tag{3.8}$$

and

$$\mathrm{E}\sup_{f^*\in\mathcal{F}}\left|S(f^*;\lambda)\frac{\partial\nabla_t(f^*;\lambda)}{\partial f}+\frac{\partial S(f^*;\lambda)}{\partial f}\cdot\nabla_t(f^*;\lambda)\right|<\frac{1-|\beta|}{|\alpha|}. \qquad (3.9)$$

Condition (3.9) is intuitive and reveals two interesting cases. If a constant scaling is used, $S(f_t(f_1,\theta);\lambda)\equiv\bar{S}$, a maximal SE region is obtained if the parameterization h is such that $\nabla_t(f_t(f_1,\theta);\lambda)=\zeta_1(u_t;\lambda)$ does not depend on $f_t(f_1,\theta)$. The condition then reduces to $|\beta|<1$ as both partial derivatives in (3.9) are equal to zero. Similarly, the SE region is maximal if the parameterization h yields a separable score $\nabla_t(f;\lambda)=\eta(f;\lambda)\zeta_1(u_t;\lambda)$ and the scaling function is $S(f)=1/\eta(f;\lambda)$. In this case, $s_t(f_t(f_1,\theta);\lambda)$ does not depend on $f_t(f_1,\theta)$ and the condition also reduces to $|\beta|<1$. We illustrate both features further in the examples in the next section.

4. Examples

To illustrate how the conditions formulated in Section 3 can be implemented for relevant empirical models, we consider a number of examples for time-varying conditional volatility, time-varying conditional expectation, and time-varying conditional tail index. The results in this section establish SE properties for a range of volatility and point process GAS models suggested in earlier work, for which the dynamic properties have so far not been characterized.

4.1. Example 1: Volatility dynamics

The case of GAS driven volatility models embeds a wide class of GARCH models. It includes new robust volatility models such as the Student's t based GAS volatility model of Creal et al. [4] and the Beta-t-GARCH model of Harvey and Chakravarty [16]. This class can be even extended to models for positively valued random variables such as dynamic duration and intensity models. It also includes robust Gamma-Weibull mixture models for duration data as proposed in Koopman et al. [20].

We can formulate this class by considering a special case of the GAS model with observation equation (2.6) to obtain the GAS scale model

$$y_t=h(f_t(f_1,\theta))u_t\ , \qquad f_{t+1}(f_1,\theta)=\omega+\alpha s_t(f_t(f_1,\theta);\lambda)+\beta f_t(f_1,\theta), \quad (4.1)$$

where $\{u_t\}$ is independently identically distributed with $u_t\sim p_{u,\lambda}$ and the function h is smooth. We typically have $\mathrm{E}[u_t]=0$ in a volatility model and $\mathrm{E}[u_t]=1$ in a duration or intensity model. Then

$$\begin{aligned}s_t(f_t(f_1,\theta);\lambda)&=S(f_t(f_1,\theta);\lambda)\cdot\nabla_t(f_t(f_1,\theta);\lambda)\\&=-S(f_t(f_1,\theta);\lambda)\cdot\nabla h(f_t(f_1,\theta))\cdot\left(\nabla p_{u,\lambda}(u_t)u_t+1\right), \quad (4.2)\end{aligned}$$

where $\nabla h(f_t(f_1, \theta)) = \partial \log h(f_t(f_1, \theta))/\partial f$ and $\nabla p_{u,\lambda}(u_t) = \partial \log p_{u,\lambda}(u_t)/\partial u_t$. It follows that

$$\frac{\partial s_t(f;\lambda)}{\partial f} = -\left(\frac{\partial \nabla h(f)}{\partial f} S(f;\lambda) + \frac{\partial S(f;\lambda)}{\partial f} \nabla h(f)\right)(\nabla p_{u,\lambda}(u_t) u_t + 1). \quad (4.3)$$

The result applies to, for example, the familiar GARCH model where $p_{u,\lambda}$ is the standard normal distribution and $S(f_t(f_1, \theta); \lambda) = \mathcal{I}_t(f_t(f_1, \theta); \lambda)^{-1}$. It also covers many other models, including models for volatility and duration dynamics as discussed in Section 1. The GAS scale model (4.1) with Gaussian disturbance sequence $\{u_t\}$ can be adopted to illustrate cases where the conditions of Assumption 4 do not hold and a non-degenerate SE region cannot be ensured.

4.1.1. Non-degeneracy of SE region

In case of the GAS scale model (4.1), Assumption 4 applies with

$$\begin{aligned} \eta(f_t(f_1, \theta); \lambda) &= S(f_t(f_1, \theta); \lambda) \frac{\partial \nabla h(f_t(f_1, \theta))}{\partial f} + \frac{\partial S(f_t(f_1, \theta); \lambda)}{\partial f} \nabla h(f_t(f_1, \theta)), \\ \zeta_1(u_t; \lambda) &= -\nabla p_{u,\lambda}(u_t) u_t - 1, \end{aligned} \quad (4.4)$$

The theory developed in Section 3 can be used to obtain the non-degeneracy of the SE region as long as $\eta(f_t(f_1, \theta); \lambda)$ is uniformly bounded, and $\mathrm{E}|\zeta_1(u_t; \lambda)| < \infty$.

First, we consider the GAS scale model with

$$h(f_t(f_1, \theta)) = f_t(f_1, \theta)^{1/2}, \qquad u_t \sim \mathrm{N}(0, 1), \qquad S(f; \lambda) = 1.$$

It follows that

$$\nabla_t(f; \lambda) = s_t(f; \lambda) = -\frac{1}{2} f^{-1}(1 - u_t^2), \quad \frac{\partial s_t(f;\lambda)}{\partial f} = \frac{1}{2} f^{-2}(1 - u_t^2).$$

Since $\eta(f; \lambda) = \frac{1}{2} f^{-2}$ is not bounded, the conditions of Assumption 4 are not satisfied. Therefore, we cannot ensure the existence of a non-degenerate SE region for this GAS model.

A different GAS model is obtained if we replace the assumption of unit scaling $S(f_t(f_1, \theta); \lambda) = 1$ by a scaling based on the inverse information matrix, that is

$$S(f_t(f_1, \theta); \lambda) = \mathcal{I}_t(f_t(f_1, \theta); \lambda)^{-1}.$$

We obtain,

$$s_t(f; \lambda) = f \cdot (u_t^2 - 1), \quad \frac{\partial s_t(f;\lambda)}{\partial f} = u_t^2 - 1. \quad (4.5)$$

4.1.2. SE region bounds

When making use of (4.4), the structure of Assumption 4 can be used to obtain bounds on the SE region. We specifically consider the parameterization

$h(f_t(f_1,\theta)) = f_t(f_1,\theta)^{1/2}$ which implies that $f_t(f_1,\theta)$ is the variance of y_t. If we set $S(f_t(f_1,\theta);\lambda) = \mathcal{I}_t(f_t(f_1,\theta);\lambda)^{-1}$, we obtain

$$s_t(f;\lambda) = -2f\cdot\mathcal{I}^{-1}_{p_{u,\lambda}}\cdot(\nabla p_{u,\lambda}u_t+1), \quad \frac{\partial s_t(f;\lambda)}{\partial f} = -2\mathcal{I}^{-1}_{p_{u,\lambda}}\cdot(\nabla p_{u,\lambda}u_t+1),$$

where $\mathcal{I}_{p_{u,\lambda}} = \mathrm{E}[(\nabla p_{u,\lambda}(u_t))^2u_t^2]-1$, which does not depend on f. This result is valid for models that are substantially different from the standard GARCH model, such as the Student's t GAS volatility model of Creal et al. [4] and the Generalized Hyperbolic GAS volatility model of Zhang et al. [30]. These models have dynamic volatility properties that are clearly different from those of the GARCH model. In particular, they correct the volatility dynamics for the fat-tailedness and possible skewness of u_t. The GAS volatility model for a Student's t distribution with λ degrees of freedom can serve as an example. Its dynamic equation for the volatility $f_t(f_1,\theta)$ is given by

$$f_{t+1}(f_1,\theta) = \omega+\beta f_t(f_1,\theta)+\alpha s_t(f_t(f_1,\theta);\lambda), \tag{4.6}$$

$$s_t(f_t(f_1,\theta);\lambda) = (1+3\lambda^{-1})\cdot\left(w_t(f_t(f_1,\theta);\lambda)y_t^2 - f_t(f_1,\theta)\right), \tag{4.7}$$

$$w_t(f_t(f_1,\theta);\lambda) = \frac{1+\lambda^{-1}}{1+\lambda^{-1}y_t^2/f_t(f_1,\theta)} = \frac{1+\lambda^{-1}}{1+\lambda^{-1}u_t^2}. \tag{4.8}$$

The weight w_t ensures that large values of y_t have a smaller impact on future values of $f_t(f_1,\theta)$; see Creal et al. [4] and Harvey [15] for more details. To ensure positivity of the variance $f_t(f_1,\theta)$ at all times, it follows directly from (4.6) that we require $\beta > (1+3\lambda^{-1})\alpha > 0$. If $\lambda^{-1} = 0$, these restrictions collapse to the standard restrictions for the GARCH model.[2] Using these restrictions, we obtain the simplification

$$\begin{aligned}\mathrm{E}\sup_{f^*}\left|\beta+\alpha\frac{\partial s_t(f^*)}{\partial f}\right| &= \mathrm{E}\left|\beta-(1+3\lambda^{-1})\alpha+(1+3\lambda^{-1})\alpha\frac{(1+\lambda^{-1})u_t^2}{1+\lambda^{-1}u_t^2}\right| \\ &= \beta-(1+3\lambda^{-1})\alpha+(1+3\lambda^{-1})\alpha\ \mathrm{E}\left[\frac{(1+\lambda^{-1})u_t^2}{1+\lambda^{-1}u_t^2}\right] \\ &= \beta-(1+3\lambda^{-1})\alpha+(1+3\lambda^{-1})\alpha = \beta. \end{aligned} \tag{4.9}$$

Hence analytical bounds are immediately given by $\beta < 1$ subject to conditions that ensure positivity of $f_t(f_1,\theta)$ for all t, i.e., $\beta > (1+3\lambda^{-1})\alpha > 0$. Note that $\beta < 1$ for the GAS parameter β coincides with the familiar condition $\alpha^*+\beta^* < 1$ for the GARCH parameters α^* and β^*; see also footnote 2.

Figure 2 presents the SE regions obtained by the numerical evaluation of condition (3.4) and the analytical derivations based on condition (4.9) for different values of λ. The (linearly) upward sloping lower bound of the SE region follows from the condition $f_t(f_1,\theta) > 0$ and is given by the relation $\beta = (1+3\lambda^{-1})\alpha$.

[2] The parameters α and β of the GAS model coincide to the familiar α^* and $(\alpha^*+\beta^*)$ parameters, respectively, for the standard GARCH model as in Bollerslev [1]. Hence the restrictions $\beta > \alpha > 0$ for the GAS parameters are the same as $\alpha^*, \beta^* > 0$ in the standard GARCH model.

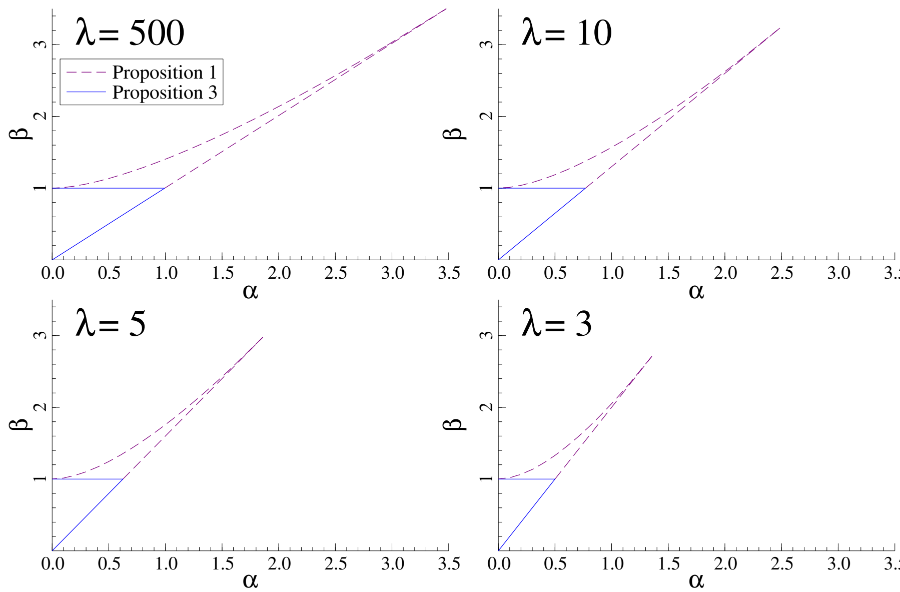

FIG 2. *Stationarity and ergodicity regions for the Student's t GAS volatility model for different values of λ obtained by Proposition 1 under numerical evaluation of Assumption 2 (dashed line) and the region derived analytically in Proposition 3 (solid line).*

Finally, the curved region is obtained by numerical integration of equation (3.4). The difference between the curved SE region and the solid triangle is a consequence of Jensen's inequality when going from sufficient condition (3.4) to (3.5).

4.1.3. Maximal SE regions

We have discussed in Section 4.2.3 that for particular choices of the parameterization h and scale S, we can obtain the maximal SE region. In case of model (4.1), we can set $h(f_t(f_1, \theta)) = \exp(f_t(f_1, \theta))$ and $S(f_t(f_1, \theta); \lambda) = 1$, which implies that we model log volatility with unit scaling. This parameterization can be convenient to ensure positivity of the variance without imposing parameter restrictions on α or β. This model has been used in, for example, Janus et al. [18] and Harvey [15], and its multivariate counterparts in Creal et al. [4] and Zhang et al. [30]. It is easily shown that $s_t(f_t(f_1, \theta); \lambda)$ does not depend on $f_t(f_1, \theta)$ for this specification. As a result, $\partial s_t(f; \lambda)/\partial f = 0$ and we obtain the maximal SE region $|\beta| < 1$ without any restrictions on α.

We conclude this example by investigating the influence of the scaling function S. In particular, we consider the GAS model (4.1) with $S(f_t(f_1, \theta); \lambda) = \mathcal{I}_t(f_t(f_1, \theta); \lambda)^{-1/2}$ for some arbitrary parameterization $h(f_t(f_1, \theta))$. We then have

$$s_t(f; \lambda) = -\mathcal{I}_{p_{u,\lambda}}^{-1/2} \cdot (\nabla p_{u,\lambda} u_t + 1),$$

which does not depend on $f_t(f_1,\theta)$ and hence yields the maximum SE region $|\beta|<1$ for any arbitrary parameterization $h(f_t(f_1,\theta))$ and a square root inverse information matrix scaling.[3] This is a specific case where the effect of the parameterization $h(f_t(f_1,\theta))$ on the size and shape of the SE region vanishes for a specific choice of the scaling function S.

4.2. Example 2: Conditional expectation models

GAS models of the conditional expectation embed a large class of linear and nonlinear state-space models of the conditional mean as analyzed for example in Hamilton [14] and Durbin and Koopman (2012, Chapter 2) [8], with applications in Stock and Watson [26]. In addition, it gives rise to various linear and nonlinear autoregressive moving average (ARMA) model specifications for the dynamics of observed data as covered for example in Granger and Terasvirta [13].

For concreteness, consider a GAS model for the time-varying conditional expectation,

$$y_t = h\big(f_t(f_1,\theta)\big)+u_t, \qquad f_{t+1}(f_1,\theta) = \omega+\alpha s_t\big(f_t(f_1,\theta);\lambda\big)+\beta f_t(f_1,\theta), \tag{4.10}$$

where $\{u_t\}$ is an independently identically distributed random variable with $u_t \sim p_u(\cdot;\lambda) = p_{u,\lambda}$, such that $\mathrm{E}[u_t]=0$, and hence,

$$\mathrm{E}[y_t|y_{t-1},y_{t-2},\ldots] = \mathrm{E}\big[y_t|f_t(f_1,\theta)\big] = h\big(f_t(f_1,\theta)\big).$$

We assume that h is a continuous, smooth function of the time-varying parameter $f_t(f_1,\theta)$. Given the assumption on the pdf of u_t, the pdf of y_t takes the form $p_y(y_t|f_t(f_1,\theta)) = p_{u,\lambda}(y_t - h(f_t(f_1,\theta)))$, such that

$$\frac{\partial}{\partial f}\log p_y(y_t|f_t(f_1,\theta)) = -h'(f_t(f_1,\theta))\frac{\partial}{\partial u_t}\log p_{u,\lambda}(u_t). \tag{4.11}$$

As a result, we obtain the following specification for the GAS step s_t,

$$s_t(f_t(f_1,\theta);\lambda) = -S(f_t(f_1,\theta);\lambda)\cdot h'(f_t(f_1,\theta))\cdot \nabla p_{u,\lambda}(u_t), \tag{4.12}$$

with

$$\begin{aligned}\frac{\partial s_t(f_t(f_1,\theta);\lambda)}{\partial f} = &-\big[h''(f_t(f_1,\theta))S(f_t(f_1,\theta);\lambda)\\ &+ h'(f_t(f_1,\theta))S'(f_t(f_1,\theta);\lambda)\big]\cdot\nabla p_{u,\lambda}(u_t),\end{aligned} \tag{4.13}$$

where $h'(f_t(f_1,\theta))$ and $h''(f_t(f_1,\theta))$ denote the first and second order derivatives of h, respectively, evaluated at $f_t(f_1,\theta)$, $S'(f_t(f_1,\theta);\lambda)$ denotes the derivative of the scale function $S(\cdot;\lambda)$, evaluated at $f_t(f_1,\theta)$, and $\nabla p_{u,\lambda}(u_t)$ denotes the score

[3] The SE region may be smaller than $|\beta|<1$ for any α if parameter restrictions apply to α and β, for example, to ensure positivity of $f_t(f_1,\theta)$ for the Student's t GAS model with $h(f_t(f_1,\theta)) = f_t(f_1,\theta)^{1/2}$.

of the error density w.r.t. u_t, that is $\nabla p_{u,\lambda}(u_t) = \partial \log p_{u,\lambda}(u_t)/\partial u_t$. The score function $\nabla p_{u,\lambda}(u_t)$ depends on the static parameter λ, but not on the dynamic parameter $f_t(f_1, \theta)$.

This GAS model can also give rise to nonlinear ARMA dynamics for $\{y_t\}_{t\in\mathbb{Z}}$ as given by

$$y_{t+1} = h^*(y_t, u_t; \theta) + u_{t+1} \; \forall \; t \in \mathbb{Z},$$

where

$$h^*(y_t, u_t; \theta) = h\Big(\omega + \alpha - S(h^{-1}(y_t - u_t); \lambda) \\ \times h'\big(h^{-1}(y_t - u_t)\big) \cdot \nabla p_{u,\lambda}(u_t) + \beta\big(h^{-1}(y_t - u_t)\big)\Big),$$

and h^{-1} denotes the inverse of h. A linear ARMA specification is only obtained if h and S are linear functions.

4.2.1. Non-degeneracy of SE region

Since Assumptions 1 and 3 hold, the non-degeneracy of the SE region of this GAS model can be established by bounding $\mathrm{E}\sup_{f^*\in\mathcal{F}} |\partial s_t(f^*; \lambda)/\partial f|$ and by appealing to Lemma 1 and Propositions 1 and 2.[4] The moment bound on the score derivative $\mathrm{E}\sup_{f^*\in\mathcal{F}} |\partial s_t(f^*; \lambda)/\partial f|$ can in turn be obtained by ensuring that Assumption 4 holds and then by appealing to Proposition 3.

To see that the current GAS model fits the conditions in Assumption 4, we set $\zeta_2(u_t; \lambda) = 0$, $\zeta_1(u_t; \lambda) = -\nabla p_{u,\lambda}(u_t)$, and

$$\eta(f_t(f_1, \theta); \lambda) = h''(f_t(f_1, \theta))S(f_t(f_1, \theta); \lambda) + h'(f_t(f_1, \theta))S'(f_t(f_1, \theta); \lambda).$$

Hence, by Propositions 2 and 3, we need to show that $h'' \cdot S + h' \cdot S'$ is uniformly bounded and $\mathrm{E}|\nabla p_{u,\lambda}| < \infty$. For example, when we consider the case of unit scaling, $S(f_t(f_1, \theta); \lambda) = 1$ with independently identically distributed Gaussian errors $u_t \sim \mathrm{N}(0, \sigma^2)$, it follows that $\lambda = \sigma^2$. In this case we obtain $s_t(f_t(f_1, \theta); \lambda) = h'(f_t(f_1, \theta))\sigma^{-2}u_t$, and

$$\frac{\partial s_t(f_t(f_1, \theta); \sigma^2)}{\partial f} = h''(f_t(f_1, \theta))\sigma^{-2}u_t.$$

The non-degeneracy of the SE region is then obtained by bounding

$$\mathrm{E}\sup_{f^*\in\mathcal{F}} \left|\frac{\partial s_t(f^*; \sigma^2)}{\partial f}\right| = \mathrm{E}\sup_{f^*\in\mathcal{F}} \left|h''(f^*)\sigma^{-2}u_t\right|.$$

In the context of Assumption 4, set $\zeta_2(u_t; \sigma^2) = 0$, $\eta(f; \sigma^2) = h''(f)$, $\zeta_1(u_t; \sigma^2) = \sigma^{-2}u_t$. Then the conditions of Proposition 3 are satisfied on a non-degenerate SE region if $h''(f)$ is uniformly bounded, $0 \leq \sigma^2 < \infty$ and $\mathbb{E}|u_t| < \infty$.

[4]For Assumption 1 we have that $\{\omega + \beta f + s_t(f; \lambda), t \in \mathbb{Z}\}$ is a stationary and ergodic (SE) sequence of Lipschitz maps due to the assumption of an independently identically distributed sequence $\{u_t\}$ and the subsequent conditions imposed on h and the distribution $p_{u,\lambda}$.

If a more elaborate scaling function S is used, then non-degeneracy is available for a larger class of nonlinear link functions h. In particular, by Assumption 4 and Proposition 3 the results can be extended to unbounded link functions h with unbounded second derivatives, since the relevant boundedness condition on $\eta(f;\sigma^2)$ must hold for $\eta(f;\sigma^2) = h''(f)S(f;\lambda) + h'(f)S'(f;\lambda)$ instead.

4.2.2. SE region bounds

To provide bounds on the SE region, we use Proposition 2, which in this case reduces to considering

$$\mathrm{E}\sup_{f^*\in\mathcal{F}}\left|\beta+\alpha\sigma^{-2}u_t h''(f^*)\right| \leq |\beta| + |\alpha|\sigma^{-2}\sup_{f^*\in\mathcal{F}}|h''(f^*)|\ \mathrm{E}|u_t|. \tag{4.14}$$

As a concrete example, consider the logistic link function $h(f) = (1+\exp(-f))^{-1}$ with Gaussian errors $u_t \sim \mathrm{N}(0,\sigma^2)$, and unit scaling $S \equiv 1$. Then $h''(f) = -(e^{2f} - e^f)/(e^{3f} + 3e^{2f} + 3e^f + 1)$, and hence equation (4.14) reduces to

$$|\beta| + \frac{|\alpha|\sqrt{3}}{18\sigma}\cdot\mathrm{E}\left|\frac{u_t}{\sigma}\right| = |\beta| + \frac{|\alpha|\sqrt{6}}{18\sigma\sqrt{\pi}} \approx |\beta| + 0.076776|\alpha|/\sigma. \tag{4.15}$$

This yields the sufficient SE region

$$|\beta| < 1 - \left(|\alpha|\sqrt{6}\right)\Big/\left(18\sigma\sqrt{\pi}\right). \tag{4.16}$$

For $\sigma^2 = 1$, Figure 3 plots the regions obtained by numerical evaluation of the Bougerol condition in Assumption 2, and the analytic bound obtained in (4.15).

4.2.3. Maximal SE regions

When we consider a GAS model for (4.10) and scale the scores $\nabla_t(f_t(f_1,\theta);\lambda)$ by the square root of the inverse information matrix, that is

$$\mathcal{I}_t(f_t(f_1,\theta);\lambda) = -\mathrm{E}\left[\nabla_t(f_t(f_1,\theta);\lambda)^2\right] = \sigma^{-2}h'(f_t(f_1,\theta))^2,$$

we have

$$\begin{aligned} s_t(f_t(f_1,\theta);\lambda) &= S(f_t(f_1,\theta);\lambda)\cdot\nabla_t(f_t(f_1,\theta);\lambda) \\ &= \mathcal{I}_t(f_t(f_1,\theta);\lambda)^{-1/2}\cdot\nabla_t(f_t(f_1,\theta);\lambda) \ = \ \sigma^{-1}u_t, \end{aligned}$$

which does not depend on $f_t(f_1,\theta)$. As a result, we obtain the maximal SE region characterized by $|\beta| < 1$ for this model.

The maximal SE region can also be obtained if we let the time-varying parameter be the mean, rather than the transformed mean of y_t, that is $y_t = f_t(f_1,\theta) + u_t$. The GAS model for this parameterization has $h(f) = f$ and it follows that $s_t(f_t(f_1,\theta);\lambda) = S(f_t(f_1,\theta);\lambda)u_t/\sigma^2$. Therefore, as long as the scale $S(f_t(f_1,\theta);\lambda)$ does not depend on $f_t(f_1,\theta)$, we obtain the maximal SE region. This includes all cases where $S(f_t(f_1,\theta);\lambda)$ is a power of $\mathcal{I}_t(f_t(f_1,\theta);\lambda)$.

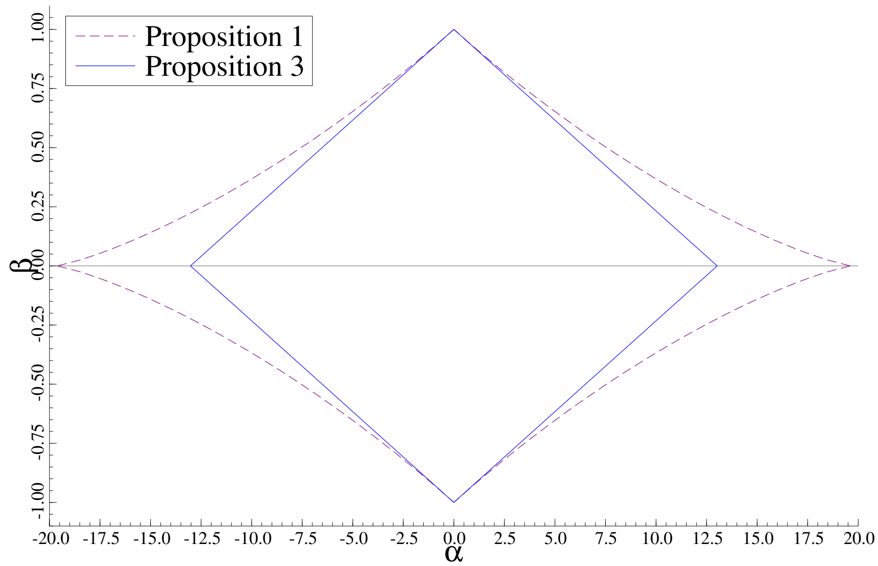

FIG 3. *Stationarity and ergodicity regions for the dynamic logistic regression model derived in Proposition 1 under numerical evaluation of Assumption 2 (dashed line) and the region derived analytically in Proposition 3 (solid line).*

4.3. Example 3: Higher-order moments

Our final example consists of a model with time-varying higher-order moments. In particular, we consider a model where the tail index $f_t(f_1,\theta)$ of a Pareto distribution is time-varying. Consider the density

$$p_y(y_t|f_t(f_1,\theta)) = f_t(f_1,\theta)^{-1} y_t^{-(1+f_t(f_1,\theta)^{-1})}, \qquad y_t > 1, \tag{4.17}$$

where $h(f_t(f_1,\theta)) = f_t(f_1,\theta) > 0$ is the tail index. The model is a special case of equation (2.6) and implies that the data is generated by

$$g(f_t(f_1,\theta), u_t) = (1-u_t)^{-f_t(f_1,\theta)}, \tag{4.18}$$

where $u_t \in (0,1)$ is a standard uniform random variable. The equivalence of the two model representations can be shown by inverting the cumulative distribution function corresponding to (4.17). The score function is given by

$$\nabla_t(f_t(f_1,\theta);\lambda) = f_t(f_1,\theta)^{-2}(\log(y_t) - f_t(f_1,\theta)) = f_t(f_1,\theta)^{-1} \cdot (-\log(1-u_t) - 1), \tag{4.19}$$

where $-\log(1-u_t)$ has a standard exponential distribution with unit mean. The information matrix is given by $\mathcal{I}_t(f_t(f_1,\theta);\lambda) = f_t(f_1,\theta)^{-2}$.

For a GAS model with unit scaling $S(f_t(f_1,\theta);\lambda) = 1$, we cannot ensure the existence of a non-degenerate SE region since $\nabla_t(f_t(f_1,\theta);\lambda)$ is unbounded in $f_t(f_1,\theta)$ for fixed u_t. For a GAS model with inverse information matrix scaling

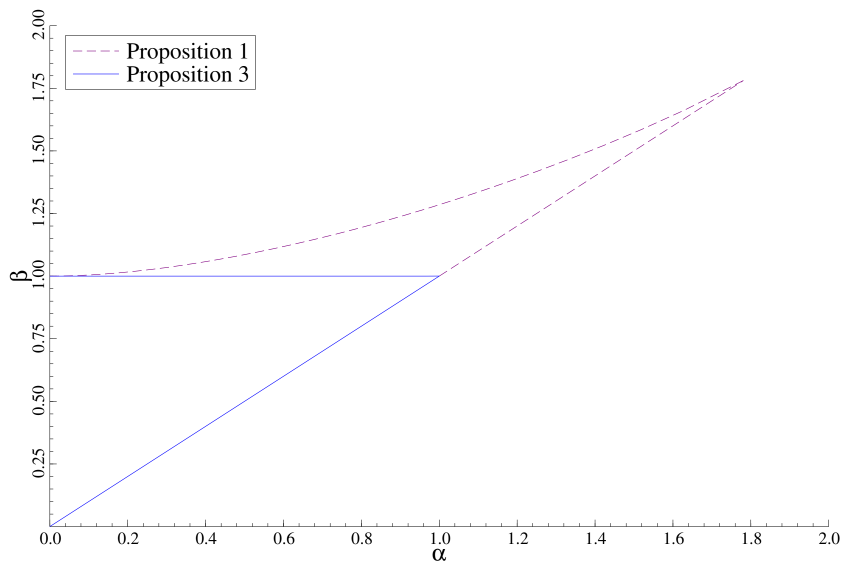

Fig 4. *Stationarity and ergodicity region for the dynamic tail index model* (4.17) *derived in Proposition 1 under numerical evaluation of Assumption 2 (dashed line) and the region derived analytically in Proposition 3 (solid line).*

$S(f_t(f_1,\theta);\lambda) = \mathcal{I}_t(f_t(f_1,\theta);\lambda)^{-1}$, $s_t(f_t(f_1,\theta);\lambda)$ is linear in $f_t(f_1,\theta)$. Therefore, its derivative does not depend on $f_t(f_1,\theta)$ and we can easily obtain the bound for the SE region by appealing to Propositions 2 and 3. The result is presented in Figure 4, where we impose the restriction $\beta > \alpha > 0$ to ensure that the tail index $f_t(f_1,\theta)$ always remains positive.

An interesting feature of our current approach is that sometimes we can facilitate the derivation of the SE region by a transformation of variables rather than by a transformation of parameters. For example, consider the GAS model for $\log(y_t)$ rather than for y_t. The Jacobian of this transformation does not depend on $f_t(f_1,\theta)$ and therefore does not influence the GAS dynamics for $f_t(f_1,\theta)$. In particular, using (4.18) we recognize that $\log(y_t)$ has an exponential distribution with mean $f_t(f_1,\theta)$. Therefore, we can consider the model specification

$$\log(y_t) = f_t(f_1,\theta)\cdot u_t^*,$$

where u_t^* is a standard exponentially distributed random variable with unit mean. This reduces the derivation of the SE region for model (4.17) to that for model (4.1). Based on this relation, the SE regions take a similar form as those in Section 4.1. This similarity also holds when we consider a GAS model with inverse square root information matrix scaling $S(f_t(f_1,\theta);\lambda) = \mathcal{I}_t(f_t(f_1,\theta);\lambda)^{-1/2}$. In this case $s_t(f_t(f_1,\theta);\lambda)$ does not depend on $f_t(f_1,\theta)$ and hence we obtain the maximal SE region $|\beta| < 1$. The same result holds if we parameterize the log tail index rather than the tail index itself and consider unit scaling $S(f_t(f_1,\theta);\lambda) = 1$.

5. Concluding remarks

In this paper we have derived conditions characterizing the stationarity and ergodicity (SE) regions for a general class of observation driven dynamic parameter models which are referred to as Generalized Autoregressive Score (GAS) models. The GAS model has a likelihood function that is analytically tractable. Given the flexibility of the GAS framework, new dynamic models of empirical interest are easily formulated. However, the dynamic specification for most GAS models is highly nonlinear. This complicates our understanding of the dynamic properties of the model.

Different formulations of the conditions for SE may be relevant for different GAS model formulations. Illustrations are provided for GAS models of time-varying means, variances, and tail shapes, whose dynamic SE properties have not been characterized in earlier work. The examples are empirically relevant and include GAS models for volatility and duration dynamics under fat-tailed distributions.

Given the current results, three obvious extensions emerge. First, it appears useful to apply our results to a proof of consistency and asymptotic normality for the maximum likelihood estimator of a class of univariate GAS models. The characterization of the SE region is a key step in obtaining laws of large numbers and central limit theorems that are required in the proof of such results. Second, it is interesting to extend our current results to the multivariate context. Third, it is interesting to use the generality of the stochastic recurrence approach to characterize the SE regions of mixed models for continuous and discrete data, such as the mixed measurement dynamic factor GAS models of Creal et al. [6]. We leave such extensions for future work.

Appendix: Proofs

Proof of Proposition 1. By Assumptions 1–3 and Lemma 1, $\{f_t(\theta)\}_{t\in\mathbb{Z}}$ is an SE sequence. By continuity of h, $\{h(f_t(\theta))\}_{t\in\mathbb{Z}}$ is a measurable sequence (w.r.t. the Borel σ-algebra). This sequence is trivially stationary. Ergodicity follows by Proposition 4.3 of Krengel (1985, p.26) [21]. Together with $\{u_t\}$ being SE (Assumption 3), it follows that $\{(u_t, h(f_t(\theta)))\}_{t\in\mathbb{Z}}$ is a stationary and ergodic vector sequence. By the same argument, continuity of g_λ ensures measurability of $y_t = g_\lambda(h(f_t(\theta)), u_t)$ and hence that $\{y_t\} = \{g_\lambda(h(f_t(\theta)), u_t)\}_{t\in\mathbb{Z}}$ is also SE. $\square$

Proof of Proposition 2. For every map $\phi_t(\cdot;\theta) : \mathcal{F} \to \mathbb{R}$, define

$$H(\phi_t(\cdot;\theta)) = \sup_{f,f'} \frac{|\phi_t(f;\theta) - \phi_t(f';\theta)|}{|f - f'|},$$

and note that,

$$\mathrm{E}\left[\log \sup_{f,f'} H(\phi_0^{(r)}(\cdot;\theta))\right]$$

$$
\begin{aligned}
&= \mathrm{E}\left[\log \sup_{f,f'} \frac{|\phi_0^{(r)}(f;\theta) - \phi_0^{(r)}(f';\theta)|}{|f-f'|}\right] \\
&= \mathrm{E}\left[\log \sup_{f,f'} \frac{|\phi_0(\cdot;\theta) \circ \cdots \circ \phi_{1-r}(f;\theta) - \phi_0(\cdot;\theta) \circ \cdots \circ \phi_{1-r}(f';\theta)|}{|f-f'|}\right] \\
&\leq \mathrm{E}\left[\log \prod_{i=1}^{r} \sup_{f,f'} \frac{|\phi_{1-i}(f;\theta) - \phi_{1-i}(f';\theta)|}{|f-f'|}\right] \\
&\leq \sum_{i=1}^{r} \mathrm{E}\left[\log \sup_{f,f'} \frac{|\phi_{1-i}(f;\theta) - \phi_{1-i}(f';\theta)|}{|f-f'|}\right],
\end{aligned}
$$

since for every collection of Lipschitz maps $\phi_0(\cdot;\theta), \ldots, \phi_{1-r}(\cdot;\theta)$ with $H(\phi_i(\cdot;\theta)) < \infty$ it holds that $H(\phi_0(\cdot;\theta) \circ \cdots \circ \phi_{1-r}(\cdot;\theta)) \leq \prod_{i=1}^{r} H(\phi_{1-r}(\cdot;\theta))$. Hence, it follows that,

$$
\begin{aligned}
&\mathrm{E}\left[\log \sup_{f,f'} \frac{|\phi_{1-i}(f;\theta) - \phi_{1-i}(f';\theta)|}{|f-f'|}\right] < 0 \; \forall i \\
&\quad \Rightarrow \quad \sum_{i=1}^{r} \mathrm{E}\left[\log \sup_{f,f'} \frac{|\phi_{1-i}(f;\theta) - \phi_{1-i}(f';\theta)|}{|f-f'|}\right] < 0 \\
&\quad \Rightarrow \quad \mathrm{E}\left[\log \sup_{f,f'} \frac{|\phi_0^{(r)}(f;\theta) - \phi_0^{(r)}(f';\theta)|}{|f-f'|}\right] < 0.
\end{aligned}
$$

We can thus focus on the condition, $\mathrm{E}[\log H(\phi_t(\cdot;\theta))] < 0$ for all $t \in \mathbb{Z}$. By Jensen's inequality,

$$
\mathrm{E}\left[\log \sup_{f,f'} \frac{|\phi_t(f;\theta) - \phi_t(f';\theta)|}{|f-f'|}\right] \leq \log \mathrm{E}\left[\sup_{f,f'} \frac{|\phi_t(f;\theta) - \phi_t(f';\theta)|}{|f-f'|}\right],
$$

such that we have the sufficient condition

$$
\mathrm{E}\left[\sup_{f,f'} \frac{|\phi_t(f;\theta) - \phi_t(f';\theta)|}{|f-f'|}\right] < 1. \tag{A1}
$$

The assumed a.s. continuous differentiability of s_t in f implies the a.s. continuous differentiability of ϕ_t in f. The exact Taylor series expansion on the realized ϕ_t states that for every $(f, f') \in \mathbb{R}^2$, $\exists f^* \in [f, f']$ such that

$$
\begin{aligned}
\phi_t(f;\theta) = \phi_t(f';\theta) + &\frac{\partial \phi_t(f^*;\theta)}{\partial f}(f - f') \\
&\Leftrightarrow \quad |\phi_t(f;\theta) - \phi_t(f';\theta)| = \left|\frac{\partial \phi_t(f^*;\theta)}{\partial f}\right| |f - f'| \\
&\Leftrightarrow \quad \frac{|\phi_t(f;\theta) - \phi_t(f';\theta)|}{|f-f'|} = \left|\frac{\partial \phi_t(f^*;\theta)}{\partial f}\right|.
\end{aligned}
$$

Now, since this holds for every pair (f, f'), then,

$$\sup_{f,f'} \frac{|\phi_t(f;\theta) - \phi_t(f';\theta)|}{|f - f'|} \quad \leq \quad \sup_{f^*} \left| \frac{\partial \phi_t(f^*;\theta)}{\partial f} \right|,$$

and hence

$$\mathrm{E} \sup_{f,f'} \frac{|\phi_t(f;\theta) - \phi_t(f';\theta)|}{|f - f'|} \quad \leq \quad \mathrm{E} \sup_{f^*} \left| \frac{\partial \phi_t(f^*;\theta)}{\partial f} \right|.$$

As a result, (A1) is implied by $\mathrm{E} \sup_{f^*} |\partial \phi_t(f^*;\theta)/\partial f| < 1$. Finally, since $\partial \phi_t(f^*;\theta)/\partial f = \beta + \alpha \cdot \partial s_t(f^*;\lambda)/\partial f$, we have that

$$\mathrm{E} \sup_{f^*} \left| \frac{\partial \phi_t(f^*;\theta)}{\partial f} \right| < 1 \quad \Leftrightarrow \quad \mathrm{E} \sup_{f^*} \left| \beta + \alpha \frac{\partial s_t(f^*;\lambda)}{\partial f} \right| < 1.$$

By norm sub-additivity, we have

$$\mathrm{E} \sup_{f^*} \left| \beta + \alpha \frac{\partial s_t(f^*;\lambda)}{\partial f} \right| \quad < \quad |\beta| + |\alpha| \cdot \mathrm{E} \sup_{f^*} \left| \frac{\partial s_t(f^*;\lambda)}{\partial f} \right|,$$

which yields the desired condition $\mathrm{E} \sup_{f^*} |\partial s_t(f^*;\lambda)/\partial f| < (1-|\beta|)/|\alpha|$. Uniformity in t follows directly from the i.i.d. nature of u_t and implies the boundedness condition at $t = 0$. □

Proof of Proposition 3. For every $\lambda \in \Lambda$, the i.i.d. nature of $\{u_t\}_{t\in\mathbb{Z}}$ and the a.s. continuous differentiability of $s_t(f;\lambda)$ in f imply that $\{\phi_t(\cdot;\theta)\}$ is SE as imposed in Assumption 1. This follows directly from the fact that $\phi_t(f;\theta) = \omega + \alpha s_t(f,\lambda) + \beta f$, $s_t(f,\lambda) = s(u_t, f, \lambda)$, and that the continuity of s in u_t implies that $\{s(u_t, \cdot, \lambda)\}$ is i.i.d. and hence also SE. Next, note that $\mathrm{E}[\log^+ |s_0(f;\lambda) - f|] < \infty$ for some $f \in \mathcal{F}$ for the process defined in (3.2) implies the condition $\mathrm{E}[\log^+ |\phi_0(f;\theta) - f|] < \infty$ for some $f \in \mathcal{F}$ in Assumption 1. To see this, we note that it follows from Lemma 2.2 in Straumann and Mikosch [27] that for all $(\omega, \alpha, \beta) \in \mathbb{R}^3$

$$\begin{aligned}
\mathrm{E} \log^+ |\phi_0(f;\theta) - f| \; &\leq \; \mathrm{E} \log^+ |\omega + \alpha s_0(f;\lambda) + \beta f - f| \\
&\leq \; 2\log 2 + \mathrm{E} \log^+ |\alpha(s_0(f;\lambda) - f)| + \mathrm{E} \log^+ |\omega + \beta f + (\alpha - 1) f| \\
&\leq \; 2\log 2 + \log^+ |\alpha| + \mathrm{E} \log^+ |s_0(f;\lambda) - f| + \log^+ |\omega + \beta f + (\alpha - 1) f|,
\end{aligned}$$

such that $\mathrm{E}[\log^+ |\phi_0(f;\lambda) - f|] < \infty$ for some $f \in \mathcal{F}$ is implied by $\mathrm{E} \log^+ |s_0(f;\lambda) - f| < \infty$ for some $f \in \mathcal{F}$. Now, continuous differentiability of $s_t(f;\lambda)$ in f implies by Proposition 2 and Lemma 1 that $\{f_t(f_1, \theta)\}_{t\in\mathbb{N}}$ as generated by (3.1) converges e.a.s. to the unique SE solution $\{f_t(\theta)\}_{t\in\mathbb{Z}}$ for every $(\omega, \alpha, \beta, \lambda)$ on a non-degenerate set

$$\left\{ (\omega, \alpha, \beta, \lambda) \in \mathbb{R} \times \mathbb{R} \times (-1, 1) \times \Lambda \; : \; |\alpha| < \frac{1 - |\beta|}{\mathrm{E} \sup_{f^*} |\partial s_t(f^*;\lambda)/\partial f|} \right\}$$

if $\mathrm{E}\sup_{f^*}|\partial s_t(f^*;\lambda)/\partial f| < \infty$. This bound is obtained under Assumption 4 with $\bar{\eta}(\lambda) < \infty$, $\bar{\zeta}_1(\lambda) < \infty$ and $\bar{\zeta}_2(\lambda) < \infty$ through norm-subadditivity since, $\forall\ t \in \mathbb{N}$,

$$\begin{aligned}\mathrm{E}\sup_{f\in\mathcal{F}}|\partial s_t(f;\lambda)/\partial f| &\leq \mathrm{E}\sup_{f\in\mathcal{F}}|\eta(f;\lambda)\zeta_1(u_t;\lambda)+\zeta_2(u_t;\lambda)| \\ &\leq \sup_{f\in\mathcal{F}}|\eta(f;\lambda)|\mathrm{E}\,|\zeta_1(u_t;\lambda)|+\mathrm{E}\,|\zeta_2(u_t;\lambda)| \\ &\leq\ \bar{\eta}(\lambda)\bar{\zeta}_1(\lambda)+\bar{\zeta}_2(\lambda).\end{aligned}$$

As a result the non-degenerate set takes the form

$$\Theta := \Big\{(\omega,\alpha,\beta,\lambda)\in\mathbb{R}\times\mathbb{R}\times(-1,1)\times\Lambda\ :\ |\alpha| < \frac{1-|\beta|}{\bar{\eta}(\lambda)\times\bar{\zeta}_1(\lambda)+\bar{\zeta}_2(\lambda)}\Big\}.$$

By continuity of h, g_λ, and the i.i.d. nature of $\{u_t\}_{t\in\mathbb{Z}}$, we obtain that $\{y_t\}_{t\in\mathbb{Z}}$ as defined in (2.6) is an SE sequence. □

References

[1] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 31*(3), 307–327. MR0853051

[2] Bougerol, P. (1993). Kalman filtering with random coefficients and contractions. *SIAM Journal on Control and Optimization 31*(4), 942–959. MR1227540

[3] Cox, D. R. (1981). Statistical analysis of time series: some recent developments. *Scandinavian Journal of Statistics 8*, 93–115. MR0623586

[4] Creal, D., Koopman, S. J., and Lucas, A. (2011). A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations. *Journal of Business and Economic Statistics 29*(4), 552–563. MR2879242

[5] Creal, D., Koopman, S. J., and Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics 28*(5), 777–795. MR3083971

[6] Creal, D., Schwaab, B., Koopman, S. J., and Lucas, A. (2013). Observation driven mixed-measurement dynamic factor models. *Review of Economics and Statistics*, forthcoming.

[7] Diaconis, P. and Freedman, D. (1999). Iterated random functions. *SIAM review*, 45–76. MR1669737

[8] Durbin, J. and Koopman, S. J. (2012). *Time Series Analysis by State Space Methods* (2nd ed.). Oxford University Press. MR3014996

[9] Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflations. *Econometrica 50*, 987–1008. MR0666121

[10] Engle, R. F. (2002). New frontiers for ARCH models. *Journal of Applied Econometrics 17*(5), 425–446.

[11] Engle, R. F. and Russell, J. R. (1998). Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica 66*, 1127–1162. MR1639411

[12] Francq, C. and Zakoian, J.-M. (2013). Inference in non stationary asymmetric GARCH models. *Annals of Statistics 41*(4), 1970–1998. MR3127855

[13] Granger, C. W. J. and Terasvirta, T. (1993). *Modelling Non-Linear Economic Relationships*. Oxford University Press.

[14] Hamilton, J. D. (1986). State-space models. In R. F. Engle and D. McFadden (Eds.), *Handbook of Econometrics*, Volume 4 of *Handbook of Econometrics*, Chapter 50, pp. 3039–3080. Elsevier. MR1315985

[15] Harvey, A. C. (2013). *Dynamic Models for Volatility and Heavy Tails*. Cambridge University Press. MR3135260

[16] Harvey, A. C. and Chakravarty, T. (2008). Beta-t-(E)GARCH. *University of Cambridge, Faculty of Economics, Working paper CWPE 08340*.

[17] Harvey, A. C. and Luati, A. (2014). Filtering with heavy tails. *Journal of the American Statistical Association*, forthcoming.

[18] Janus, P., Koopman, S. J., and Lucas, A. (2011). Long memory dynamics for multivariate dependence under heavy tails. *Tinbergen Institute Discussion Papers 11-175/DSF28*.

[19] Jensen, S. T. and Rahbek, A. (2004). Asymptotic inference for nonstationary garch. *Econometric Theory 20*(6), 1203–1226. MR2101955

[20] Koopman, S. J., Lucas, A., and Scharth, M. (2012). Predicting time-varying parameters with parameter-driven and observation-driven models. *Tinbergen Institute Discussion Papers 12-020/4*.

[21] Krengel, U. (1985). *Ergodic Theorems*. De Gruyter studies in Mathematics, Berlin. MR0797411

[22] Lucas, A., Schwaab, B., and Zhang, X. (2014). Conditional euro area sovereign default risk. *Journal of Business and Economic Statistics 32*(2), 271–284. MR3207839

[23] Nelson, D. B. (1990). Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory 6*, 318–334. MR1085577

[24] Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: a new approach. *Econometrica 59*, 347–370. MR1097532

[25] Rydberg, T. H. and Shephard, N. (2003). Dynamics of trade-by-trade price movements: decomposition and models. *Journal of Financial Econometrics 1*(1), 2.

[26] Stock, J. H. and Watson, M. W. (2006). Why has u.s. inflation become harder to forecast? NBER Working Papers 12324, National Bureau of Economic Research, Inc.

[27] Straumann, D. and Mikosch, T. (2006). Quasi-maximum-likelihood estimation in conditionally heteroeskedastic time series: A stochastic recurrence equations approach. *The Annals of Statistics 34*(5), 2449–2495. MR2291507

[28] Wintenberger, O. (2013). Continuous invertibility and stable QML estimation of the EGARCH(1, 1) model. *Scandinavian Journal of Statistics 40*(4), 846–867. MR3145121

[29] Wu, W. and Shao, X. (2004). Limit theorems for iterated random functions. *Journal of Applied Probability 41*(2), 425–436. MR2052582

[30] Zhang, X., Creal, D., Koopman, S. J., and Lucas, A. (2011). Modeling dynamic volatilities and correlations under skewness and fat tails. *Tinbergen Institute Discussion Papers 11-078/DSF22*.