

TI 2012-059/4

Tinbergen Institute Discussion Paper



Stationarity and Ergodicity of Univariate Generalized Autoregressive Score Processes

*Francisco Blasques*¹

Siem Jan Koopman^{1,2}

André Lucas^{1,2,3}

¹ Faculty of Economics and Business Administration, VU University Amsterdam;

² Tinbergen Institute;

³ Duisenberg school of finance.

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and VU University Amsterdam.

More TI discussion papers can be downloaded at <http://www.tinbergen.nl>

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 1600

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Duisenberg school of finance is a collaboration of the Dutch financial sector and universities, with the ambition to support innovative research and offer top quality academic education in core areas of finance.

DSF research papers can be downloaded at: <http://www.dsf.nl/>

Duisenberg school of finance
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 8579

Stationarity and Ergodicity of Univariate Generalized Autoregressive Score Processes*

Francisco Blasques^{a,b}, Siem Jan Koopman^{b,c}, and Andre Lucas^{a,c,d}

(a) Department of Finance, VU University Amsterdam

(b) Department of Econometrics, VU University Amsterdam

(c) Tinbergen Institute

(d) Duisenberg school of finance

June 21, 2012

Abstract

We characterize the dynamic properties of Generalized Autoregressive Score (GAS) processes by identifying regions of the parameter space that imply stationarity and ergodicity. We show how these regions are affected by the choice of parameterization and scaling, which are key features of GAS models compared to other observation driven models. The Dudley entropy integral is used to ensure the non-degeneracy of such regions. Furthermore, we show how to obtain bounds for these regions in models for time-varying means, variances, or higher-order moments.

JEL codes: C13, C22, C58;

Keywords: Dudley integral, Durations, Higher-order models, Nonlinear dynamics, Time-varying parameters, Volatility.

*We would like to thank Richard Davis and Anders Rahbek for useful searching discussions. Blasques and Lucas thank the Dutch Science Foundation (NWO) for financial support. Corresponding author: Francisco Blasques, VU University Amsterdam, FEWEB/FIN, de Boelelaan 1105, 1081 HV Amsterdam, The Netherlands. Email: f.blasques@vu.nl.

1 Introduction

The distinction between observation driven and parameter driven time series models is established in Cox (1981). The class of Generalized Autoregressive Score (GAS) models as introduced by Creal, Koopman, and Lucas (2011, 2012) is a new class of observation driven time-varying parameter models. It encompasses well-known observation driven time series models, including the autoregressive conditional heteroskedasticity (ARCH) model of Engle (1982), the generalized ARCH (GARCH) model of Bollerslev (1986), the exponential GARCH (EGARCH) model of Nelson (1991), the autoregressive conditional duration (ACD) model of Engle and Russell (1998), the multiplicative error model (MEM) of Engle (2002), the autoregressive conditional multinomial (ACM) model of Rydberg and Shephard (2003), and many related models. In addition, the GAS modeling framework gives rise to new time-varying parameter models. Examples are the multivariate volatility and correlation models for skewed and fat-tailed distributions with time-varying, possibly fractionally-integrated, parameters as specified in Creal, Koopman, and Lucas (2011), Zhang, Creal, Koopman, and Lucas (2011), and Janus, Koopman, and Lucas (2011). Other examples are the observation driven mixed measurement dynamic factor models of Creal, Schwaab, Koopman, and Lucas (2011) and the fat-tailed mixture models for duration data as proposed in Koopman, Lucas, and Scharth (2012). The latter paper also demonstrates that in terms of forecasting accuracy, GAS models perform similar to and sometimes even better than their state space or parameter driven counterparts for a range of data generating processes.

The flexibility and generality of GAS models make them interesting objects for further study. Except for some well-known special cases of the GAS class such as the ARCH and GARCH models, however, little is known about the general stochastic behavior of GAS processes. In this paper we fill this gap 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 is important, as one of the main advantages of observation driven models such as GAS models is the availability of an analytic expression for the likelihood function. The likelihood can be obtained in closed form using a standard prediction error decomposition. The consistency and asymptotic normality of the resulting maximum likelihood (ML) estimator typically depend on the stationarity and ergodicity of the underlying time series process. This is the focal point of the current paper.

Our approach builds on the stationarity and ergodicity (SE) conditions as formulated in Straumann and Mikosch (2006) for general stochastic recurrence equations. We first show that GAS models can be cast in this form. We then derive sufficient conditions for the supremum Lipschitz constants of the (stochastic) recurrence relations that need to be bounded in expectation. 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.

In our main result we use the Dudley (1967) integral to characterize the non-degeneracy of the SE region of the parameter space. Dudley (1967) has pioneered the use of entropy conditions to obtain maximal inequalities. These conditions have proved instrumental in non-parametric statistics and have played a fundamental role in the theory of empirical processes. In econometrics, Andrews (1997) and Chen (2007) provide thorough reviews of the relevant literature. Our characterization of SE regions using the Dudley integral is valid even in the presence of complex nonlinear dynamics. The Dudley integral reveals the existence of a trade-off between the complexity of the function space of derivatives for the dynamic equations associated with the time-varying parameters on the one hand, and the size of the SE region on the other hand. In particular, function spaces of lower empirical entropy allow for larger SE regions. We apply our results to a range of interesting GAS models for which SE regions have not been characterized before. They include for example volatility and duration models with different distributions, parameterizations, and scalings. The applications reveal how the conditions for stationarity and ergodicity as derived here can be validated in practice.

The remainder of this paper is organised 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. Section 5 concludes. 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,

$$p_y(y_t|h(f_t); \lambda), \quad (1)$$

for all $t \in \mathbb{Z}$, where f_t represents a scalar time-varying parameter, $h : \mathbb{R} \rightarrow \mathbb{R}$ is a link function, and $\lambda \in \Lambda$ is a vector of time-invariant parameters. This specification includes models with a time-varying mean or a time-varying volatility via the appropriate choice of f_t and h . For example, in a time-varying volatility model for a Student's t distribution, λ contains the degrees of freedom parameter, while $h(f_t)$ denotes the time-varying variance. The model can be extended to allow for exogenous variables, lagged endogenous variables, and lagged values of f_t in the conditioning set; see Creal, Koopman, and Lucas (2011, 2012).

The Generalized Autoregressive Score (GAS) model specifies a dynamic process for the parameter f_t and is given by

$$f_{t+1} = \omega + \alpha s_t(f_t; \lambda) + \beta f_t, \quad (2)$$

$$s_t(f_t; \lambda) = S(f_t; \lambda) \cdot \nabla_t(f_t; \lambda), \quad (3)$$

$$\nabla_t(f_t; \lambda) = \partial \log p_y(y_t|f_t; \lambda) / \partial f|_{f=f_t}, \quad (4)$$

where ω , α , and β are time-invariant parameters, $S(f_t; \lambda)$ is a univariate scaling factor for the score $\nabla_t(f_t; \lambda)$ of the conditional density of the observations (1), and \log denotes the natural logarithm. The current GAS model specification has one lag of $s_t(f_t; \lambda)$ and one lag of f_t in the right-hand side of (2). The inclusion of more lags for f_t or $s_t(f_t; \lambda)$ is straightforward; see Creal, Koopman, and Lucas (2012) 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) is the definition of $s_t(f_t; \lambda)$ as the scaled score of the conditional density of the observations (1) with respect to the time-varying parameter f_t . The intuition for this is straightforward: at time t we improve the local fit of the model as measured by the log density $\log p_y(y_t|f_t; \lambda)$. We do so by taking a scaled step in the steepest ascent direction of the model's fit at time t as measured by the log conditional density. Since $s_t(f_t; \lambda)$ is a function of past data and parameters alone, the GAS model can be classified as an observation driven model which is defined by Cox (1981). The equations (1) to (4) encompass a large set of familiar time series models. For example, if

p_y is the normal density, f_t is the time-varying variance, and h is the identity function, we obtain the standard GARCH model; 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 not been studied before.

Each choice for the scaling function $S = S(f_t; \lambda)$ in (3) gives rise to a new GAS model. Intuitive choices for S may be related to the local curvature of the score as measured by the inverse information matrix, for example

$$S(f_t; \lambda) = (\mathcal{I}_t(f_t; \lambda))^{-a}, \quad (5)$$

where

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

and where a is typically taken as 0, 1, or 1/2. 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 $h(f_t) = f_t$ with f_t representing the variance. Alternatively, we can opt for modelling the log variance instead by setting $h(f_t) = \exp(f_t)$ with f_t representing the log variance. The latter specification can have the advantage that the variance itself is always positive by construction, even if f_t becomes negative.

To provide further structure to the probability density function in (1) we let it be implicitly defined by the following observation equation,

$$y_t = g_\lambda(h(f_t), u_t) \quad \forall t \in \mathbb{Z}, \quad (6)$$

where for all $\lambda \in \Lambda$, $g_\lambda : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ is a function and $\{u_t\}$ is an independently identically distributed sequence with $u_t \sim p_{u,\lambda}(u_t)$. This structure covers many cases of empirical interest. For example, a time-varying volatility model is obtained by setting $y_t = f_t^{1/2} \cdot u_t$, with f_t denoting the time-varying variance and u_t being, for example, normally or Student's t distributed. Many other models are contained in (6) as well by letting g_λ be the inverse distribution function corresponding to (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 degrees of freedom.

The stochastic properties of $\{y_t\}$ 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. For this general modelling framework with the notation as established, we can 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 Straumann and Mikosch (2006) for general stochastic recurrence equations; see also Diaconis and Freedman (1999) and Wu and Shao (2004). In particular, we define subsets of $\mathcal{H} \times \mathcal{P}_u \times \mathcal{G} \times \mathcal{S} \times \Theta$ that render $\{y_t\}$ stationary and ergodic (SE). Throughout, we assume $s_t(f; \lambda)$ to be almost surely (a.s.) continuously differentiable in f . 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.

A stochastic recurrence equation for f_t takes the form

$$f_{t+1} = \phi_t(f_t; \theta) \quad \forall t \in \mathbb{Z}, \tag{7}$$

where ϕ_t is a random function. This clearly embeds the GAS model in (2) by setting

$$\phi_t(f_t; \theta) = \omega + \alpha s_t(f_t; \lambda) + \beta f_t, \tag{8}$$

with every y_t in the expression for $s_t(f_t; \lambda)$ replaced by $y_t = g_\lambda(h(f_t), u_t)$ from equation (6). Sufficient conditions for the existence of an SE sequence $\{f_t\}$ are given by the following

assumptions, where $\mathcal{F} \subseteq \mathbb{R}$ denotes the domain of f_t .

Assumption 1. For every $\theta \in \Theta$, $\{\phi_t(\cdot; \theta)\}_{t \in \mathbb{Z}}$ is a stationary and ergodic (SE) sequence of Lipschitz maps $\phi_t(\cdot; \theta) : \mathbb{R} \rightarrow \mathbb{R}$ satisfying $\mathbb{E} [\log^+ \|\phi_0(f; \theta) - f\|] < \infty$ for some $f \in \mathcal{F}$.

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

$$\mathbb{E} \log \left[\sup_{(f, f') \in \mathcal{F} \times \mathcal{F}: f \neq f'} \frac{\|\phi_0(f; \theta) - \phi_0(f'; \theta)\|}{\|f - f'\|} \right] < \infty, \quad (9)$$

and

$$\mathbb{E} \log \left[\sup_{(f, f') \in \mathcal{F} \times \mathcal{F}: f \neq f'} \frac{\|\phi_0^{(r)}(f; \theta) - \phi_0^{(r)}(f'; \theta)\|}{\|f - f'\|} \right] < 0, \quad (10)$$

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

The proof of the following lemma can be found in Straumann and Mikosch (2006).¹

Lemma 1. Let Assumptions 1 and 2 hold and let $\{f_t\}$ be generated by (7). Then for every $\theta \in \Theta$, $\{f_t\}$ is an SE random sequence.

The SE properties for $\{y_t\}$ follow directly from those of $\{f_t\}$ given the model's structure in (6). This is stated in the following assumption and proposition.

Assumption 3. The sequence $\{u_t\}_{t \in \mathbb{Z}}$ in (6) consists of independently identically distributed random variables and the function g_λ is continuous for every $\lambda \in \Lambda$.

Proposition 1. Let Assumptions 1–3 hold and let $\{f_t\}$ be generated by (7). Let the link function $h : \mathbb{R} \rightarrow \mathbb{R}$ be continuous. Then $\{y_t\}$ 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 , such as stationarity and ergodicity, but for our current expositional purposes the assumption of an independently identically distributed u_t suffices.

Numerical evaluation of condition (10) in Assumption 2 is complicated by the fact that we have to take the expectation of the supremum of a possibly highly non-linear function. Mistakes can then easily be made if the numerical algorithm for computing the supremum fails to find the global supremum for every $\theta \in \Theta$ and every $u_t \in \mathbb{R}$. To

¹The condition in (10) is used to ensure stationarity and ergodicity of a sequence in \mathbb{N} that starts at $t = 0$, while the relevant stochastic sequence in (7) is for $t \in \mathbb{Z}$. We therefore need to show that the sequence in \mathbb{N} approximates the sequence in \mathbb{Z} and hence that both have the same properties. We refer to Straumann and Mikosch (2006) for a formal proof.

prevent such potential mistakes, we provide a further analytical characterization of the conditions under which the SE region can be proven to be non-degenerate. The following proposition is helpful in this respect.

Proposition 2. *Condition (10) in Assumption 2 is implied by,*

$$\mathbb{E} \sup_{f^* \in \mathcal{F}} \left| \beta + \alpha \frac{\partial s_t(f^*; \lambda)}{\partial f} \right| < 1, \quad (11)$$

which in turn is implied by

$$\mathbb{E} \sup_{f^* \in \mathcal{F}} \left| \frac{\partial s_t(f^*; \lambda)}{\partial f} \right| < \frac{1 - |\beta|}{|\alpha|}. \quad (12)$$

In applications, both conditions (11) and (12) play important roles to characterize the SE region; we illustrate their roles in Section 4.

From condition (12) it follows directly that the SE region:

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

Given (i), we focus our discussion entirely on the size of the SE region in the (α, β) -plane. For a given value of λ , we obtain the maximum SE region if

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

In this case, only condition (ii) above is binding, while $\alpha^- = -\infty$, and $\alpha^+ = \infty$, such that condition (iii) 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,

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

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

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

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 (12), 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 (12) 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 . 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, Koopman, and Lucas (2011). In particular, their SE properties and conditions have as yet not been fully investigated. The results for an $s_t(\cdot; \lambda)$ that is affine in f_t are summarized in the following corollary.

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

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

3.2 Non-Degeneracy and bounds for SE regions

The two steps in analyzing the dynamic properties of GAS models are (i) to ensure that the SE region is non-degenerate; and (ii) to determine the SE region's actual size and shape. As shown in Section 3.1, the existence and size of a non-degenerate SE region both depend on the value of $E \sup_{f^* \in \mathcal{F}} |\partial s_t(f^*; \lambda) / \partial f|$. In what follows, we obtain meaningful upper bounds for this expectation given the link function h , family of distributions p_u , and scale function S .

In Proposition 2, the variable f takes the role of a parameter. As a result, $\partial s_t(f^*; \lambda) / \partial f$ in Assumption 3 is a random variable only due to its dependence on u_t . Define the function $\xi : \mathbb{R} \times \mathcal{F} \times \Lambda \rightarrow \mathbb{R}$ as

$$\xi(u_t, f; \lambda) \equiv \xi_f(u_t; \lambda) \equiv \xi_{f,\lambda}(u_t) \equiv \frac{\partial s_t(f; \lambda)}{\partial f}.$$

For every $(f, \lambda) \in \mathcal{F} \times \Lambda$, the function $\xi_{f,\lambda} : \mathbb{R} \rightarrow \mathbb{R}$ is then a real-valued function, while $\xi_{f,\lambda}(u_t)$ is a real-valued random variable. In what follows, we focus on the stochastic process $X_{\xi_{f,\lambda}}$, where

$$\{\xi_{f,\lambda}(u_t)\}_{f \in \mathcal{F}} \equiv \{\xi_{f,\lambda}(u_t)\}_{\xi_{f,\lambda} \in \Xi_\lambda} \equiv \{X_{\xi_{f,\lambda}}\}_{\xi_{f,\lambda} \in \Xi_\lambda},$$

and the process is indexed by the functions $\xi_{f,\lambda} \in \Xi_\lambda$ with

$$\Xi_\lambda \equiv \{\xi_{f,\lambda} \mid f \in \mathcal{F}\},$$

for fixed $\lambda \in \Lambda$. Our main quantity of interest then becomes

$$\mathbb{E} \sup_{\xi_{f,\lambda} \in \Xi_\lambda} |X_{\xi_{f,\lambda}}| = \mathbb{E} \sup_{f \in \mathcal{F}} |\xi_{f,\lambda}(u_t)|. \quad (13)$$

The switch in notation from the right-hand side of (13) to the left-hand side may at first seem unnecessarily complicated. This switch, however, allows us to make direct use of the *Dudley integral* (Dudley (1967)), which provides a maximal inequality for separable *sub-Gaussian* stochastic processes. The Dudley integral relates the upper bound on $\mathbb{E} \sup_{\xi_{f,\lambda}} |X_{\xi_{f,\lambda}}|$ to the *entropy* of the set that indexes the stochastic process under an appropriate pseudo-metric²; see, amongst others, van der Vaart and Wellner (1996, Section 2.2) and Korosok (2009, Chapters 2 and 3). Interestingly, while the index set $\mathcal{F} = \mathbb{R}$ has infinite entropy under pseudo-metrics of interest, the index set Ξ_λ can easily have finite entropy as long as appropriate conditions hold for the functions $\xi_{f,\lambda}$ of u_t generated by \mathcal{F} . Concrete examples are given in Section 4.

Definition 1. A stochastic process $\{X_{\xi_{f,\lambda}}\}_{\xi_{f,\lambda} \in \Xi_\lambda}$ is separable with respect to (w.r.t.) a countable dense subset $\Xi_\lambda^* \subseteq \Xi_\lambda$ if its sample paths are a.s. Ξ_λ^* -separable.³

Definition 2. A stochastic process $\{X_{\xi_{f,\lambda}}\}_{\xi_{f,\lambda} \in \Xi_\lambda}$ is said to be sub-Gaussian w.r.t. the pseudo-metric $d_{\Xi_\lambda} : \Xi_\lambda \times \Xi_\lambda \rightarrow \mathbb{R}_0^+$ on its index set if and only if

$$\mathbb{P}(|X_{\xi_{f,\lambda}} - X_{\xi_{f',\lambda}}| > x) \leq 2e^{-\frac{1}{2}x^2/d_{\Xi_\lambda}^2(\xi_{f,\lambda}, \xi_{f',\lambda})},$$

²A pseudo-metric d_{Ξ_λ} satisfies all the usual properties of a metric except that $d_{\Xi_\lambda}(\xi_{f,\lambda}, \xi_{f',\lambda}) = 0 \not\Rightarrow \xi_{f,\lambda} = \xi_{f',\lambda}$ for $\xi_{f,\lambda}, \xi_{f',\lambda} \in \Xi_\lambda$. A typical example of a pseudo-metric is the standard deviation of the difference between two random variables.

³A function ι on Ξ_λ is Ξ_λ^* -separable if $\forall \xi_{f,\lambda} \in \Xi_\lambda \exists \{\xi_{f,\lambda,i}^*\}_{i \in \mathbb{N}} \in \Xi_\lambda^*$ such that $\xi_{f,\lambda,i}^* \rightarrow \xi_{f,\lambda}$ as $i \rightarrow \infty$ and $\iota(\xi_{f,\lambda,i}^*) \rightarrow \iota(\xi_{f,\lambda})$. Separability is not restrictive (possibly under compactification of \mathbb{R}) since every process taking values on a compact metric space with a separable metric index space admits a separable modification (it admits a separable version with the same finite-dimensional distributions).

for every $(\xi_{f,\lambda}, \xi_{f',\lambda}) \in \Xi_\lambda \times \Xi_\lambda$.⁴

Definition 3. Let $(\Xi_\lambda, d_{\Xi_\lambda})$ be a pseudo-metric space. Given $\varepsilon > 0$, $\xi_{f_1,\lambda}, \dots, \xi_{f_n,\lambda} \in \Xi_\lambda$ are ε -separated if $d_{\Xi_\lambda}(\xi_{f_i,\lambda}, \xi_{f_j,\lambda}) > \varepsilon$ for any $i \neq j$. The ε -packing number of $(\Xi_\lambda, d_{\Xi_\lambda})$, denoted by $\mathcal{D}(\varepsilon, \Xi_\lambda, d_{\Xi_\lambda})$, is the maximal cardinality of an ε -separated set. The ε -packing-entropy of $(\Xi_\lambda, d_{\Xi_\lambda})$ is given by $\mathcal{E}_{\mathcal{D}}(\varepsilon, \Xi_\lambda, d_{\Xi_\lambda}) = \log \mathcal{D}(\varepsilon, \Xi_\lambda, d_{\Xi_\lambda})$.

As functions of ε , the numbers $\mathcal{D}(\varepsilon, \Xi_\lambda, d_{\Xi_\lambda})$ diverge as $\varepsilon \rightarrow 0$. Dudley (1967) controls the complexity of Ξ_λ through an integrability condition which ensures that $\mathcal{D}(\varepsilon, \Xi_\lambda, d_{\Xi_\lambda})$ diverges at an appropriate rate as $\varepsilon \rightarrow 0$.

Lemma 2. Let $\{X_{\xi_{f,\lambda}}, \xi_{f,\lambda} \in \Xi_\lambda\}$ be a separable sub-Gaussian stochastic process w.r.t. the pseudo-metric d_{Ξ_λ} on Ξ_λ . Then, $\exists K > 0$ such that for any given $\xi_{f_0,\lambda} \in \Xi_\lambda$,

$$\mathbb{E} \sup_{\xi_{f,\lambda}} |X_{\xi_{f,\lambda}}| \leq \mathbb{E} |X_{\xi_{f_0,\lambda}}| + K \int_0^\nu \sqrt{\mathcal{E}_{\mathcal{D}}(\varepsilon, \Xi_\lambda, d_{\Xi_\lambda})} d\varepsilon,$$

where $\nu = \text{diam}(\Xi_\lambda)$ denotes the diameter of Ξ_λ .

A finite value for $\mathbb{E} \sup_{\xi_{f,\lambda}} |X_{\xi_{f,\lambda}}|$ is thus obtained whenever $\int_0^\nu \sqrt{\mathcal{E}_{\mathcal{D}}(\varepsilon, \Xi_\lambda, d_{\Xi_\lambda})} d\varepsilon$ is bounded and $\mathbb{E} |X_{\xi_{f_0,\lambda}}|$ is bounded for some $\xi_{f_0,\lambda} \in \Xi_\lambda$. Classes of functions satisfying these bounds can be very general and are well known in Empirical Process Theory and include monotone function classes and square integrable function classes; see, for example, Lemma 3 in the Appendix.

The Dudley integral and Lemma 3 in the Appendix show that under the structure of Assumption 3 every GAS model as defined in Section 2 and characterized by (h, p_u, S) has a non-degenerate SE region as long as there exists a pseudo-metric d_{Ξ_λ} on Ξ_λ for which the derivative $\partial s_t(f; \lambda) / \partial f$ is sub-Gaussian and square integrable (with $d_{\Xi_\lambda} \lesssim \|\cdot\|_2^{p_u, \lambda}$). A particularly useful case that characterizes the SE region for a wide class of empirically relevant GAS models is given by the following assumption and proposition.

Assumption 4. The partial derivative $\partial s_t(f; \lambda) / \partial f = \xi(u_t, f; \lambda)$ is a.s. continuous and

$$|\xi(u_t, f; \lambda)| \leq |\eta(f; \lambda) \zeta_1(u_t; \lambda) + \zeta_2(u_t; \lambda)|, \quad (14)$$

$$|\xi(u_t, f; \lambda) - \xi(u_t, f'; \lambda)| \leq |\eta^*(f, f'; \lambda) \zeta^*(u_t; \lambda)|, \quad (15)$$

⁴Any Gaussian process is sub-Gaussian for the standard-deviation pseudo-metric on Ξ_λ given by $d_{\Xi_\lambda}(\xi_{f,\lambda}, \xi_{f',\lambda}) = \text{stdev}(X_{\xi_{f,\lambda}} - X_{\xi_{f',\lambda}})$.

for all $(f, f') \in \mathcal{F} \times \mathcal{F}$ where $\eta(\cdot; \lambda)$ and $\eta^*(\cdot, \cdot; \lambda)$ are bounded, $\zeta_1(\cdot; \lambda)$, $\zeta_2(\cdot; \lambda)$ and $\zeta^*(\cdot; \lambda)$ are $L_2(p_u(u_t; \lambda))$ maps, and $\zeta^*(u_t; \lambda)$ satisfies an exponential tail bound $\mathbb{P}[|\zeta^*(u_t; \lambda)| > x] \leq k_1 \exp(-k_2 x^2)$ for some $(k_1, k_2) \in \mathbb{R}_0^+ \times \mathbb{R}_0^+$.

Proposition 3. *Let $\lambda \in \Lambda$ be such that Assumption 4 holds. Furthermore, let d_{Ξ_λ} be a pseudo-metric satisfying $d_{\Xi_\lambda} \lesssim \|\cdot\|_2^{p_{u,\lambda}}$ and $d_{\Xi_\lambda}(\xi_{f,\lambda}, \xi_{f',\lambda}) \geq |\eta^*(f, f'; \lambda)|$ for every $(f, f') \in \mathcal{F} \times \mathcal{F}$. Then, $\mathbb{E} \sup_{f^*} |\partial s_t(f^*; \lambda) / \partial f| < \infty$, w.r.t. d_{Ξ_λ} .*

The proposition follows directly from Dudley's integral and the complexity of the class of square integrable functions. The boundedness of η and the $L_2(p_u(u_t; \lambda))$ integrability of ζ_1 and ζ_2 in Assumption 4 ensure the square integrability of $\xi_{f,\lambda}$ for all $f \in \mathcal{F}$. The sub-Gaussianity of the stochastic process $\{X_{\xi_{f,\lambda}}\}_{\xi_{f,\lambda} \in \Xi_\lambda} = \{\xi_{f,\lambda}(u_t)\}_{f \in \mathcal{F}}$ follows from the separable bound on the derivative's difference in (15) and the exponential tail bound on the random variable $\zeta^*(u_t; \lambda)$.

It is important to note that the exponential tail bound in Assumption 4 may appear more strict than it actually is. For example, it may seem that we exclude time-varying parameter models for fat-tailed observations. This is not the case. We show in Section 4 that the exponential tail bound is easily satisfied for a wide range of GAS models for time-varying volatility or duration and fat-tailed distributions. The key to this result is the observation that the exponential tail bound need not hold for the density itself, but for differences in the derivative of the score of the density. For example, for the GAS time-varying volatility model considered in Section 4, the score is linear in f even for a highly leptokurtic Student's t distribution. Given this linearity, we have $X_{\xi_{f,\lambda}} - X_{\xi_{f',\lambda}} = 0$ for fixed u_t and hence the exponential tail bound is trivially satisfied. The result holds for more general cases: due to the consideration of differences of derivatives of scores in Proposition 3, the exponential tail bound on $X_{\xi_{f,\lambda}}$ is satisfied for a large class of GAS models, including models for fat-tailed densities.

In some cases, the entropy results from Lemma 2 allow us to take one further step and establish not only the existence of a non-degenerate SE region, but also bounds on its size. For example, Lemma 3 in the Appendix reveals that the region satisfying (12) becomes larger if the L_2 envelope of the relevant function becomes smaller.

3.3 Maximal SE regions

We can show that a large class of GAS models exhibits a maximal SE region. Note that $s_t(f_t; \lambda) = S(f_t; \lambda) \cdot \nabla_t(f_t; \lambda)$ and rewrite (11) and (12) of Proposition 2 as

$$\mathbb{E} \log \sup_{f^* \in \mathcal{F}} \left| \beta + \alpha \left(S(f^*; \lambda) \frac{\partial \nabla_t(f^*; \lambda)}{\partial f} + \frac{\partial S(f^*; \lambda)}{\partial f} \cdot \nabla_t(f^*; \lambda) \right) \right| < 0, \quad (16)$$

and

$$\mathbb{E} \sup_{f^* \in \mathcal{F}} \left| S(f^*; \lambda) \frac{\partial \nabla_t(f^*; \lambda)}{\partial f} + \frac{\partial S(f^*; \lambda)}{\partial f} \cdot \nabla_t(f^*; \lambda) \right| < \frac{1 - |\beta|}{|\alpha|}. \quad (17)$$

Condition (17) is intuitive and reveals two interesting cases. If a constant scaling is used, $S(f_t; \lambda) \equiv \bar{S}$, a maximal SE region is obtained if the parameterization h is such that $\nabla_t(f_t; \lambda) = \zeta_1(u_t; \lambda)$ does not depend on f_t . The condition then reduces to $|\beta| < 1$ as both partial derivatives in (17) 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, $s_t(f_t; \lambda)$ does not depend on f_t and the condition also reduces to $|\beta| < 1$. The examples in Section 4 reveal that these simple conditions with maximal SE regions already cover a wide range of empirically relevant GAS models.

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 mean, time-varying volatility, and time-varying 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: mean dynamics

Consider a model for the time-varying mean,

$$y_t = h(f_t) + u_t, \quad f_{t+1} = \omega + \alpha s_t(f_t; \lambda) + \beta f_t, \quad (18)$$

where $\{u_t\}$ is an independently identically distributed random variable with $u_t \sim p_u(\cdot; \lambda) = p_{u, \lambda}$, such that $\mathbb{E}[u_t] = 0$. We assume that h is a continuous, smooth function of the time-

varying parameter f_t . Given the assumption on the pdf of u_t , the pdf of y_t takes the form $p_y(y_t|f_t) = p_u(y_t - h(f_t))$, such that

$$\frac{\partial}{\partial f} \log p_y(y_t|f_t) = -h'(f_t) \frac{\partial}{\partial u_t} \log p_{u,\lambda}(u_t). \quad (19)$$

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

$$s_t(f_t; \lambda) = -S(f_t; \lambda) \cdot h'(f_t) \cdot \nabla p_{u,\lambda}(u_t), \quad (20)$$

with

$$\frac{\partial s_t(f_t; \lambda)}{\partial f} = -[h''(f_t)S(f_t; \lambda) + h'(f_t)S'(f_t; \lambda)] \cdot \nabla p_{u,\lambda}(u_t), \quad (21)$$

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

4.1.1 Non-degeneracy of SE region

Since Assumptions 1 and 3 hold, the non-degeneracy of the SE region of this GAS model for a time-varying mean can be established by bounding $\mathbb{E} \sup_{f^* \in \mathcal{F}} |\partial s_t(f^*; \lambda) / \partial f|$ and by appealing to Lemma 1 and Propositions 1 and 2.⁵ The bound on $\mathbb{E} \sup_{f^* \in \mathcal{F}} |\partial s_t(f^*; \lambda) / \partial f|$ can in turn be obtained first by ensuring that Assumption 4 holds and second by appealing to Proposition 3.

For the current GAS model with only a time-varying mean, $s_t(f; \lambda)$ fits the conditions in Assumption 4 by setting $\zeta_2(u_t; \lambda) = 0$. It follows that

$$\eta(f_t; \lambda) = h''(f_t)S(f_t; \lambda) + h'(f_t)S'(f_t; \lambda), \quad \zeta_1(u_t; \lambda) = \zeta^*(u_t; \lambda) = -\nabla p_{u,\lambda}(u_t),$$

$$\eta^*(f, f'; \lambda) = [h''(f)S(f; \lambda) + h'(f)S'(f; \lambda)] - [h''(f')S(f'; \lambda) + h'(f')S'(f'; \lambda)]. \quad (22)$$

Hence we need to show that $h'' \cdot S + h' \cdot S'$ is bounded, $-\nabla p_{u,\lambda}$ is $L_2(p_{u,\lambda})$, and $\nabla p_{u,\lambda}(u_t)$ satisfies the exponential tail bound $\mathbb{P}(|\nabla p_{u,\lambda}(u_t)| > x) \leq k_1 \exp(-k_2 x^2)$ for some $(k_1, k_2) \in$

⁵For 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 $p_{u,\lambda}$.

$\mathbb{R}_0^+ \times \mathbb{R}_0^+$. By Propositions 2 and 3, the boundedness of $\mathbb{E} \sup_{f^* \in \mathcal{F}} |\partial s_t(f^*; \lambda) / \partial f| < \infty$ and the non-degeneracy of the SE region follow under any pseudo-metric d_{Ξ_λ} that satisfies $d_{\Xi_\lambda} \lesssim \|\cdot\|_2^{p_{u,\lambda}}$ and

$$d_{\Xi_\lambda}(\xi_{f,\lambda}, \xi_{f',\lambda}) \geq |\eta^*(f, f'; \lambda)| \quad \forall (f, f') \in \mathcal{F} \times \mathcal{F},$$

with η^* as defined in (22).

For example, when we consider the case of unit scaling, $S(f_t; \lambda) = 1$ with independently identically distributed Gaussian errors $u_t \sim \mathcal{N}(0, \sigma^2)$, it follows that $\lambda = \sigma^2$. In this case we obtain $s_t(f_t; \lambda) = h'(f_t)\sigma^{-2}u_t$, and

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

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

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

In the context of Assumption 4, set $\zeta_2(u_t; \sigma^2) = 0$, $\eta(f; \sigma^2) = h''(f)$, $\zeta_1(u_t; \sigma^2) = \zeta^*(u_t; \sigma^2) = \sigma^{-2}u_t$, and $\eta^*(f, f'; \lambda) = h''(f) - h''(f')$. Given the normality assumption, $\sigma^{-2}u_t$ naturally satisfies the exponential tail bound.⁶ The remaining conditions are thus satisfied if $h''(f)$ is bounded and $\sigma^2 < \infty$ is such that $\sigma^{-2}u_t$ is $L_2(p_{u,\lambda})$. Finally, by Proposition 3, the boundedness of $\mathbb{E} \sup_{f^* \in \mathcal{F}} |h''(f^*)\sigma^{-2}u_t|$ holds w.r.t. to the standard deviation pseudo-metric d_{Ξ_λ} , since $d_{\Xi_\lambda} \lesssim \|\cdot\|_2^{p_{u,\lambda}}$ and

$$\begin{aligned} d_{\Xi_\lambda}(\xi_{f,\lambda}, \xi_{f',\lambda}) &= \text{stdev}(X_{\xi_{f,\lambda}} - X_{\xi_{f',\lambda}}) = \text{stdev}(h''(f)\sigma^{-2}u_t - h''(f')\sigma^{-2}u_t) \\ &= |h''(f) - h''(f')|\sigma^{-1} \geq |h''(f) - h''(f')| \quad \forall (f, f') \in \mathcal{F} \times \mathcal{F}. \end{aligned}$$

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.

⁶When u_t is fat-tailed, the definition of $s_t(f_t; \lambda)$ changes and the exponential tail bound may still be satisfied. For example, this is the case for the Student's t distribution where the score is bounded in u_t for finite degrees of freedom.

4.1.2 SE region bounds

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

$$\mathbb{E} \sup_{f^* \in \mathcal{F}} |\beta + \alpha \sigma^{-2} u_t h''(f^*)|. \quad (23)$$

First consider the case $\alpha, \beta > 0$ and $u_t > 0$. In that case, we can compute the supremum as

$$\beta + \alpha \sigma^{-2} u_t \sup_{f^* \in \mathcal{F}} |h''(f^*)|. \quad (24)$$

Using the symmetry of expression (23), we obtain

$$\mathbb{E} \sup_{f^* \in \mathcal{F}} |\beta + \alpha \sigma^{-2} u_t h''(f^*)| = \beta + \alpha \sigma^{-2} \mathbb{E}|u_t| \sup_{f^* \in \mathcal{F}} |h''(f^*)|. \quad (25)$$

As a concrete example, consider the logistic link function $h(f) = (1 + \exp(-f))^{-1}$ with Gaussian errors $u_t \sim \mathcal{N}(0, \sigma^2)$, and unit scaling $S \equiv 1$. Then $h''(f) = -(e^{2f} - e^f)/(e^{3f} + 3e^{2f} + 3e^f + 1)$, and hence equation (25) reduces to

$$\beta + \frac{\alpha \sqrt{3}}{18\sigma} \cdot \frac{u_t}{\sigma} = \beta + \frac{\alpha \sqrt{6}}{18\sigma \sqrt{\pi}} \approx \beta + 0.076776 \alpha / \sigma. \quad (26)$$

Appealing again to symmetry arguments, we obtain the SE region

$$|\beta| < 1 - \left(|\alpha| \sqrt{6} \right) / (18\sigma \sqrt{\pi}). \quad (27)$$

For $\sigma^2 = 1$, Figure 1 plots the regions obtained by numerical evaluation of the Straumann-Mikosch condition in Assumption 2, and the analytic bound obtained in (26).

4.1.3 Maximal SE regions

Finally note that if we consider a GAS model for (18) where we scale the scores $\nabla_t(f_t; \lambda)$ by the square root of the inverse information matrix

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

we obtain

$$s_t(f_t; \lambda) = S(f_t; \lambda) \cdot \nabla_t(f_t; \lambda) = \mathcal{I}_t(f_t; \lambda)^{-1/2} \cdot \nabla_t(f_t; \lambda) = \sigma^{-1} u_t.$$

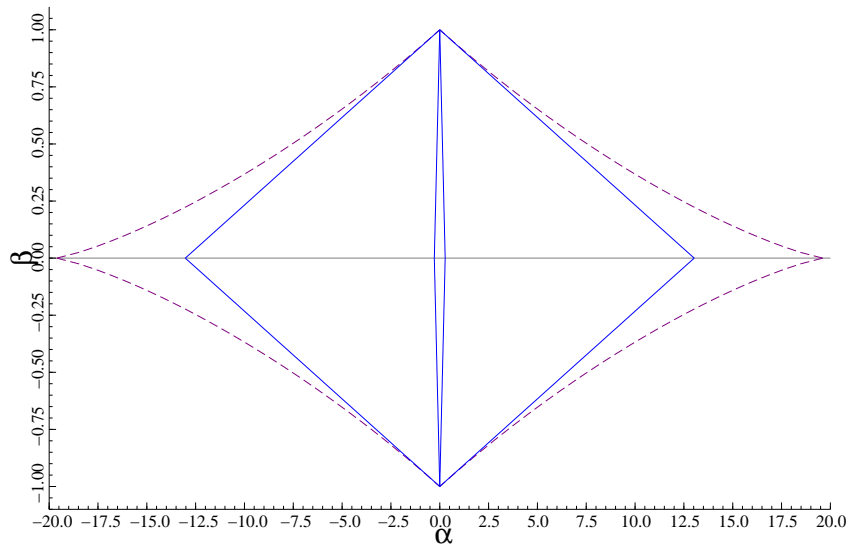


Figure 1: Stationarity and ergodicity regions for the dynamic logistic regression model

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 $y_t = f_t + u_t$. The GAS model for this parameterization has $h(f) = f$ and it follows that $s_t(f_t; \lambda) = S(f_t; \lambda)u_t/\sigma^2$. Therefore, as long as the scale $S(f_t; \lambda)$ does not depend on f_t , we obtain the maximal SE region. This includes all cases where $S(f_t; \lambda)$ is a power of $\mathcal{I}_t(f_t; \lambda)$.

4.2 Example 2: volatility dynamics

The case of volatility models is particularly interesting, as it embeds new robust volatility models such as the Student's t based GAS volatility model of Creal, Koopman, and Lucas (2011) and the Beta- t -Garch model of Harvey and Chakravarty (2008) as well as new models for positively valued random variables, such as the robust Gamma-Weibull mixture models for duration data as proposed in Koopman, Lucas, and Scharth (2012). A special case of the GAS model with observation equation (6) is the GAS scale model that we specify as

$$y_t = h(f_t)u_t, \quad f_{t+1} = \omega + \alpha s_t(f_t; \lambda) + \beta f_t, \quad (28)$$

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

intensity model. Then

$$s_t(f_t; \lambda) = S(f_t; \lambda) \cdot \nabla_t(f_t; \lambda) = -S(f_t; \lambda) \cdot \nabla h(f_t) \cdot (\nabla p_{u,\lambda}(u_t)u_t + 1), \quad (29)$$

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

$$\frac{\partial s_t(f; \lambda)}{\partial f} = - \left(\frac{\partial \nabla h(f)}{\partial f} S(f; \lambda) + \frac{\partial S(f; \lambda)}{\partial f} \nabla h(f) \right) (\nabla p_{u,\lambda}(u_t)u_t + 1). \quad (30)$$

The result applies to, for example, the familiar GARCH model where $p_{u,\lambda}$ is the standard normal distribution and $S(f_t; \lambda) = \mathcal{I}_t(f_t; \lambda)^{-1}$. It also covers many other models, including models for volatility and duration dynamics as discussed in Section 1. The GAS scale model (28) 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. In addition, Gaussian and Student's t based examples are useful to illustrate that in some cases the derivation of bounds on the SE region is easy and the Dudley integral vanishes.

4.2.1 Non-degeneracy of SE region

In case of the GAS scale model (28), Assumption 4 applies with

$$\begin{aligned} \eta(f_t; \lambda) &= S(f_t; \lambda) \frac{\partial \nabla h(f_t)}{\partial f} + \frac{\partial S(f_t; \lambda)}{\partial f} \nabla h(f_t), \\ \zeta_1(u_t; \lambda) &= \zeta^*(u_t; \lambda) = -\nabla p_{u,\lambda}(u_t)u_t - 1, \\ \eta^*(f, f'; \lambda) &= \eta(f; \lambda) - \eta(f'; \lambda). \end{aligned} \quad (31)$$

The theory developed in Section 3 can be used to obtain the non-degeneracy of the SE region as long as $\eta(f_t; \lambda)$ is bounded, and $\zeta_1(u_t; \lambda) = \zeta^*(u_t; \lambda)$ is $L_2(p_{u,\lambda})$ and satisfies the exponential tail bound.

Let us first consider the GAS scale model with

$$h(f_t) = f_t^{1/2}, \quad u_t \sim \text{N}(0, 1), \quad S(f; \lambda) = 1.$$

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 $S(f_t; \lambda) = 1$ by a scaling based on the inverse information matrix, that is $S(f_t; \lambda) = \mathcal{I}_t(f_t; \lambda)^{-1}$. We obtain,

$$s_t(f; \lambda) = f \cdot (u_t^2 - 1), \quad \frac{\partial s_t(f; \lambda)}{\partial f} = u_t^2 - 1. \quad (32)$$

In this case the non-degeneracy of the SE region can be obtained by noting that the Dudley entropy integral in Lemma 2 vanishes. This is due to the fact $\partial s_t(f; \lambda)/\partial f$ does not depend on f and, therefore, the diameter of the function space Ξ_λ is zero ($\nu = 0$) under any pseudo-metric of interest. Hence, from Lemma 2 we have $E \sup_{f^* \in \mathcal{F}} |\partial s_t(f^*; \lambda)/\partial f| \leq E|1 - u_t^2|/2 < \infty$. The non-degeneracy of the SE region can also be obtained without an appeal to the Dudley integral by observing that $s_t(f_t; \lambda)$ is linear in f_t and by relying on Corollary 1.

4.2.2 SE region bounds

When making use of (31), the structure of Assumption 4 can be used to obtain bounds on the SE region. We specifically consider the parameterization $h(f_t) = f_t^{1/2}$ which implies that f_t is the variance of y_t . If we set $S(f_t; \lambda) = \mathcal{I}_t(f_t; \lambda)^{-1}$, we obtain

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

where $\mathcal{I}_{p_{u,\lambda}} = E[(\nabla p_{u,\lambda}(u_t))^2 u_t^2] - 1$, which does not depend on f . As a result, we obtain $\nu = \text{diam}(\Xi_\lambda) = 0$ for every relevant pseudo-metric on Ξ_λ and the Dudley integral vanishes. 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, Koopman, and Lucas (2011) and the Generalized Hyperbolic GAS volatility model of Zhang, Creal, Koopman, and Lucas (2011). 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 is given by

$$f_{t+1} = \omega + \beta f_t + \alpha s_t(f_t; \lambda), \quad s_t(f_t; \lambda) = (1 + 3\lambda^{-1}) \cdot (w_t(f_t; \lambda) y_t^2 - f_t), \quad (33)$$

where

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

The weight w_t ensures that large values of y_t have a smaller impact on future values of f_t ; see Creal, Koopman, and Lucas (2011) for more details. To ensure positivity of the variance f_t at all times, it follows directly from (33) that we require $\beta > (1 + 3\lambda^{-1})\alpha > 0$. If $\lambda^{-1} = 0$, these restrictions collapse to the standard restrictions for the GARCH model⁷. Using these restrictions, we obtain the simplification

$$\begin{aligned} \mathbb{E} \sup_{f^*} \left| \beta + \alpha \frac{\partial s_t(f^*)}{\partial f} \right| &= \mathbb{E} \left| \beta - (1 + 3\lambda^{-1})\alpha + (1 + 3\lambda^{-1})\alpha \frac{(1 + \lambda^{-1})u_t^2}{1 + \lambda^{-1}u_t^2} \right| \\ &= \beta - (1 + 3\lambda^{-1})\alpha + (1 + 3\lambda^{-1})\alpha \mathbb{E} \left[\frac{(1 + \lambda^{-1})u_t^2}{1 + \lambda^{-1}u_t^2} \right] \\ &= \beta - (1 + 3\lambda^{-1})\alpha + (1 + 3\lambda^{-1})\alpha = \beta. \end{aligned} \quad (34)$$

Hence analytical bounds are immediately given by $|\beta| < 1$ subject to conditions that ensure positivity of f_t for all t .

Figure 2 presents the SE regions obtained by the numerical evaluation of condition (10) and the analytical derivations based on condition (34) for different values of λ . The (linearly) upward sloping lower bound of the SE region follows from the condition $f_t > 0$ and is given by the relation $\beta = (1 + 3\lambda^{-1})\alpha$. The horizontal line $\beta = 1$ is the upper bound of the SE region as obtained using the Dudley integral. Finally, the curved region is obtained by numerical integration of equation (10). The difference between the curved SE region and the solid triangle is a consequence of Jensen's inequality when going from sufficient condition (10) to (11).

4.2.3 Maximal SE regions

We have discussed in Section 4.1.3 that for particular choices of the parameterization h and scale S , we can obtain the maximal SE region. In case of model (28), we can set $h(f_t) = \exp(f_t)$ and $S(f_t; \lambda) = 1$, which implies that we model log volatility with unit scaling. This

⁷The 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 (1986). Hence the restrictions $\beta > \alpha > 0$ for the GAS parameter are the same as $\alpha^*, \beta^* > 0$ in the standard GARCH model.

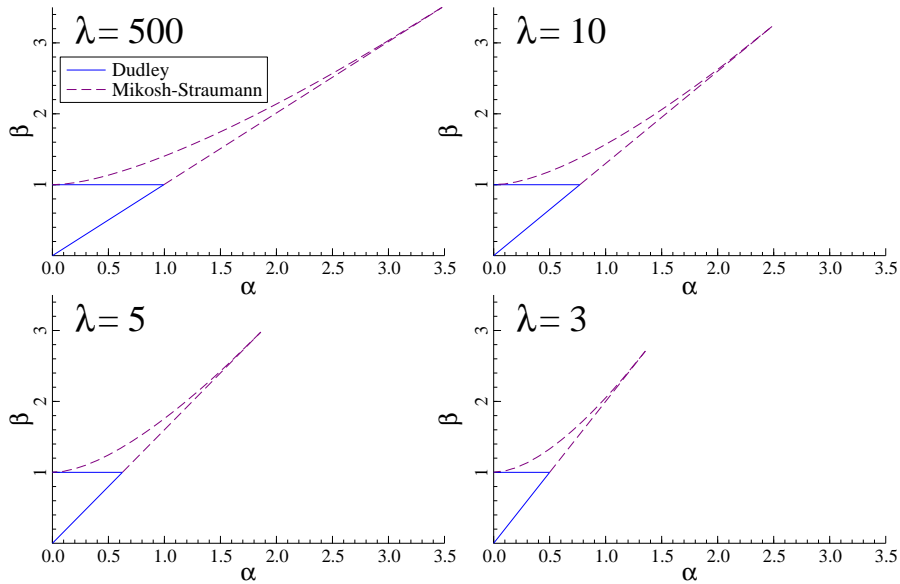


Figure 2: Stationarity and ergodicity regions for the Student's t GAS volatility model for different values of λ .

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, Koopman, and Lucas (2011), and its multivariate counterparts in Creal, Koopman, and Lucas (2011) and Zhang, Creal, Koopman, and Lucas (2011). It is easily shown that $s_t(f_t; \lambda)$ does not depend on f_t for this specification. As a result, $\partial s_t(f; \lambda) / \partial f = 0$ and we obtain the maximal SE region $|\beta| < 1$.

We conclude this example by investigating the influence of the scaling function S . In particular, we consider the GAS model (28) with $S(f_t; \lambda) = \mathcal{I}_t(f_t; \lambda)^{-1/2}$ for some arbitrary parameterization $h(f_t)$. 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 and hence yields the maximum SE region $|\beta| < 1$ for an arbitrary parameterization $h(f_t)$ and a square root inverse information matrix scaling.⁸ This is a specific case of the more general framework described in Section 3.3 where the effect of the parameterization $h(f_t)$ on the size and shape of the SE region vanishes for a specific choice of the scaling function S .

⁸The SE region may be smaller than $|\beta| < 1$ if parameter restrictions apply to α and β , for example, to ensure positivity of f_t for the Student's t GAS model with $h(f_t) = f_t^{1/2}$.

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 of a Pareto distribution is time-varying. Consider the density

$$p_y(y_t|f_t) = f_t^{-1}y_t^{-(1+f_t^{-1})}, \quad y_t > 1, \quad (35)$$

where $h(f_t) = f_t > 0$ is the tail index. The model is a special case of equation (6) and implies that the data is generated by

$$g(f_t, u_t) = (1 - u_t)^{-f_t}, \quad (36)$$

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 (35). The score function is given by

$$\nabla_t(f_t; \lambda) = f_t^{-2}(\log(y_t) - f_t) = f_t^{-1} \cdot (-\log(1 - u_t) - 1), \quad (37)$$

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

For a GAS model with unit scaling $S(f_t; \lambda) = 1$, we cannot ensure the existence of a non-degenerate SE region since $\nabla_t(f_t; \lambda)$ is unbounded in f_t for fixed u_t . For a GAS model with inverse information matrix scaling $S(f_t; \lambda) = \mathcal{I}_t(f_t; \lambda)^{-1}$, $s_t(f_t; \lambda)$ is linear in f_t . Therefore, its derivative does not depend on f_t and we can use the Dudley integral with $\nu = \text{diam}(\Xi_\lambda) = 0$ to obtain the bound for the SE region. The result is presented in Figure 3, where we impose the restriction $\beta > \alpha > 0$ to ensure that the tail index f_t 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 and therefore does not influence the GAS dynamics for f_t . In particular, using (36) we recognize that $\log(y_t)$ has an exponential distribution with mean f_t . Therefore, we can consider the model specification

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

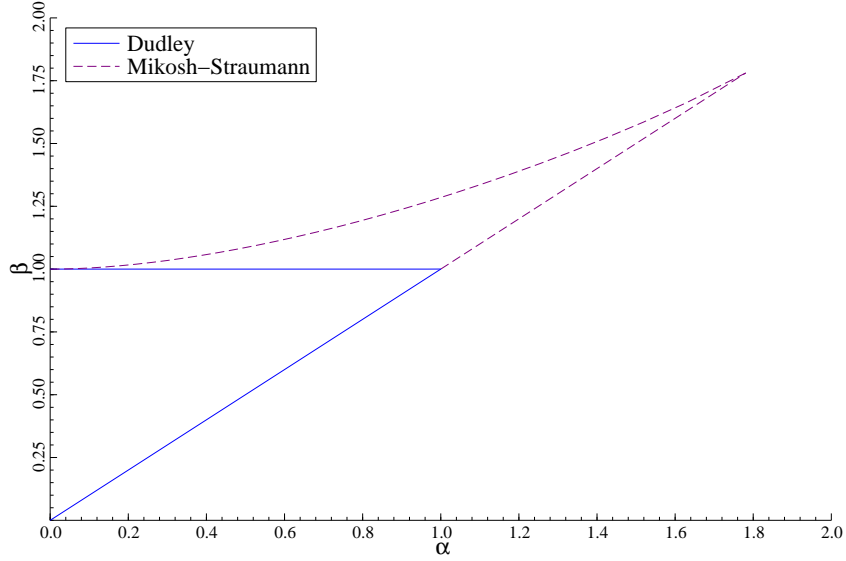


Figure 3: Stationarity and ergodicity region for the dynamic tail index model (35)

where u_t^* is a standard exponentially distributed random variable with unit mean. It reduces the derivation of the SE region for model (35) to equation (28). Based on this relation, the SE regions take a similar form as those in Section 4.2. This similarity also holds when we consider a GAS model with inverse square root information matrix scaling $S(f_t; \lambda) = \mathcal{I}_t(f_t; \lambda)^{-1/2}$. In this case $s_t(f_t; \lambda)$ does not depend on f_t 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 $S(f_t; \lambda) = 1$.

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 non-linear. This complicates our understanding of the dynamic properties of the model. We have shown that the Dudley integral provides a useful mechanism to characterize the SE region for GAS models and their dynamic processes. In particular, the Dudley integral relates the existence of non-degenerate SE regions to the complexity of the space of functions characterizing the dynamics of the time-varying parameter. The higher the complexity and empirical entropy

of the function spaces, the more difficult it is to ensure a bounded SE region.

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, Schwaab, Koopman, and Lucas (2011). We leave such extensions for future work.

References

- Andrews, D. (1997). Empirical process methods in econometrics. In R. F. Engle and D. M. Fadden (Eds.), *Handbook of Econometrics*, Volume 4, Chapter 37, pp. 2247–2294. Elsevier.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31(3), 307–327.
- Chen, X. (2007). Large sample sieve estimation of semi-nonparametric models. In R. F. Engle and D. M. Fadden (Eds.), *Handbook of Econometrics*, Volume 6b, Chapter 76, pp. 5549–5632. Elsevier.
- Cox, D. R. (1981). Statistical analysis of time series: some recent developments. *Scandinavian Journal of Statistics* 8, 93–115.
- Creal, D., S. J. Koopman, and A. Lucas (2011). A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations. *Journal of Business and Economic Statistics* 29(4), 552–563.
- Creal, D., S. J. Koopman, and A. Lucas (2012). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, forthcoming.

- Creal, D., B. Schwaab, S. J. Koopman, and A. Lucas (2011). Observation driven mixed-measurement dynamic factor models with an application to credit risk. *Tinbergen Institute Discussion Papers 11-042/DSF16*.
- Diaconis, P. and D. Freedman (1999). Iterated random functions. *SIAM review*, 45–76.
- Dudley, R. M. (1967). The sizes of compact subsets of hilbert space and continuity of gaussian processes. *Journal of Functional Analysis* 1, 290–330.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflations. *Econometrica* 50, 987–1008.
- Engle, R. F. (2002). New frontiers for ARCH models. *Journal of Applied Econometrics* 17(5), 425–446.
- Engle, R. F. and J. R. Russell (1998). Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica*, 1127–1162.
- Harvey, A. C. and T. Chakravarty (2008). Beta- t -(E)GARCH. *University of Cambridge, Faculty of Economics, Working paper CWPE 08340*.
- Janus, P., S. J. Koopman, and A. Lucas (2011). Long memory dynamics for multivariate dependence under heavy tails. *Tinbergen Institute Discussion Papers 11-175/DSF28*.
- Koopman, S. J., A. Lucas, and M. Scharth (2012). Predicting time-varying parameters with parameter-driven and observation-driven models. *Tinbergen Institute Discussion Papers 12-020/4*.
- Kosorok, M. R. (2009). Introduction to empirical processes and semiparametric inference. *International Statistical Review* 77(2), 318–318.
- Krengel, U. (1985). *Ergodic theorems*. Berlin: De Gruyter studies in Mathematics.
- Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: a new approach. *Econometrica*, 347–370.
- Rydberg, T. H. and N. Shephard (2003). Dynamics of trade-by-trade price movements: decomposition and models. *Journal of Financial Econometrics* 1(1), 2.
- Straumann, D. and T. Mikosch (2006). Quasi-maximum-likelihood estimation in conditionally heteroskedastic time series: A stochastic recurrence equations approach. *The Annals of Statistics* 34(5), 2449–2495.

van de Geer, S. (2000). *Empirical processes in M-estimation*. Cambridge University Press.

van der Vaart, A. W. and J. A. Wellner (1996). *Weak convergence and empirical processes*. Springer-Verlag, New York.

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

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

Appendix: Proofs

Proof of Proposition 1: By Assumptions 1-3 and Lemma 1, $\{f_t\}$ is an SE sequence. By continuity of h , $\{h(f_t)\}$ 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). Together with $\{u_t\}$ being SE (Assumption 3), it follows that $\{(u_t, h(f_t))\}$ is a stationary and ergodic vector sequence. By the same argument, continuity of g_λ ensures measurability of $y_t = g_\lambda(h(f_t), u_t)$ and hence that $\{y_t\} = \{g_\lambda(h(f_t), u_t)\}$ is also SE. ■

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

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

and note that,

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

since for every collection of Lipschitz maps $\phi_0(\cdot; \theta), \dots, \phi_{1-r}(\cdot; \theta)$ with $H(\phi_i(\cdot; \theta)) < \infty$ it holds that

$H(\phi_0(\cdot; \theta) \circ \dots \circ \phi_{1-r}(\cdot; \theta)) \leq \prod_{i=1}^r H(\phi_{1-r}(\cdot; \theta))$. Hence, it follows that,

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

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

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

such that we have the sufficient condition

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

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

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

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

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

and hence

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

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

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

By norm sub-additivity, we have

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

which yields the desired condition $\mathbb{E} \sup_{f^*} \|\partial s_t(f^*; \lambda) / \partial f\| < (1 - |\beta|) / |\alpha|$. \blacksquare

The following lemma illustrates the kind of entropy bounds that have been obtained for specific classes of functions; see, for example, van der Vaart and Wellner (1996) and van de Geer (2000, Lemma

3.8, p.36, and Corollary 2.6, pgs. 20 and 152) for more results. We let $d_{\Xi_\lambda} \lesssim \|\cdot\|_q^{p_{u,\lambda}}$ mean that d_{Ξ_λ} is weaker than $\|\cdot\|_q^{p_{u,\lambda}}$ and define $\|\xi_{f,\lambda}\|_q^{p_{u,\lambda}} := (\int |\xi_{f,\lambda}|^q p_{u,\lambda}(u_t) du_t)^{1/q}$.

Lemma 3. (*Square Integrable Class*) Let $\Xi_\lambda := \{\xi_{f,\lambda} : \mathbb{R} \rightarrow \mathbb{R}, \xi_{f,\lambda} \in L_2(p_{u,\lambda}), \|\xi_{f,\lambda}\|_2^{p_{u,\lambda}} \leq \delta_\lambda\}$. Then $\mathcal{E}_{\mathcal{D}}(\varepsilon, \Xi_\lambda, \|\cdot\|_2^{p_{u,\lambda}}) \leq \log(4\delta_\lambda + \varepsilon) - \log(\varepsilon) \quad \forall \varepsilon > 0$ and hence it follows that

$$\int_0^\nu \sqrt{\mathcal{E}_{\mathcal{D}}(\varepsilon, \Xi_\lambda, d_{\Xi_\lambda})} d\varepsilon \leq \nu \int_0^1 \sqrt{\log(4t_\lambda + x) - \log(x)} dx \quad \text{with} \quad t_\lambda = \frac{\delta_\lambda}{\nu}.$$

Proof of Proposition 3: By Lemma 2, the bound on $E \sup_{\xi_{f,\lambda}} |X_{\xi_{f,\lambda}}|$ is achieved if $X = \{X_{\xi_{f,\lambda}}, \xi_{f,\lambda} \in \Xi_\lambda\} = \{\partial s_t(f; \lambda) / \partial f, f \in \mathcal{F}\}$ is a separable sub-Gaussian process and the empirical entropy integral $\int_0^\nu \sqrt{\mathcal{E}_{\mathcal{D}}(\varepsilon, \Xi_\lambda, d_{\Xi_\lambda})} d\varepsilon$ is bounded. Separability of the stochastic process $X_{\xi_{f,\lambda}}$ is implied by a.s. path continuity. This is ensured in our case by a.s. continuous differentiability of s_t . Sub-Gaussianity of $X_{\xi_{f,\lambda}}$ w.r.t. d_{Ξ_λ} follows from the fact that,

$$\begin{aligned} \mathbb{P} \left[\left| X_{\xi_{f,\lambda}} - X_{\xi_{f',\lambda}} \right| > x \right] &= \mathbb{P} \left[|\xi_{f,\lambda}(u_t) - \xi_{f',\lambda}(u_t)| > x \right] \leq \mathbb{P} \left[|\eta^*(f, f'; \lambda) \zeta^*(u_t; \lambda)| > x \right] \\ &= \mathbb{P} \left[|\eta^*(f, f'; \lambda)| \cdot |\zeta^*(u_t; \lambda)| > x \right] = \mathbb{P} \left[|\zeta^*(u_t; \lambda)| > x / |\eta^*(f, f'; \lambda)| \right], \end{aligned}$$

where the inequality follows by the separable bound in (15). Using the exponential tail bound, we have,

$$\mathbb{P} \left[|\zeta^*(u_t; \lambda)| > x / |\eta^*(f, f'; \lambda)| \right] \leq k_1 \exp \left(-k_2 \frac{x^2}{\eta^*(f, f'; \lambda)^2} \right) \leq k_1 \exp \left(-k_2 \frac{x^2}{d_{\mathcal{F}}(f, f')^2} \right).$$

for any pseudo-metric $d_{\mathcal{F}}$ satisfying $d_{\mathcal{F}}(f, f') \geq |\eta^*(f, f'; \lambda)|$. Hence, sub-Gaussianity of X is obtained w.r.t. to the pseudo-metric $d_{\Xi_\lambda}(\xi_{f,\lambda}, \xi_{f',\lambda})$,

$$\mathbb{P} \left[\left| X_{\xi_{f,\lambda}} - X_{\xi_{f',\lambda}} \right| > x \right] \leq k_1 \exp \left(-k_2 \frac{x^2}{d_{\mathcal{F}}(f, f')^2} \right) \leq 2 \exp \left(-\frac{1}{2} \frac{x^2}{d_{\Xi_\lambda}(\xi_{f,\lambda}, \xi_{f',\lambda})^2} \right),$$

where $d_{\Xi_\lambda}(\xi_{f,\lambda}, \xi_{f',\lambda}) = k_3 \cdot d_{\mathcal{F}}(f, f')$ for some appropriate $k_3 > 0$.

By Lemma 3, the empirical entropy integral $\int_0^\nu \sqrt{\mathcal{E}_{\mathcal{D}}(\varepsilon, \Xi_\lambda, d_{\Xi_\lambda})} d\varepsilon$ is bounded if $\xi_{f,\lambda} \in L_2(p_{u,\lambda})$ for every $\xi_{f,\lambda} \in \Xi_\lambda$ with finite uniform empirical L_2 bound, $\sup_{\xi_{f,\lambda} \in \Xi_\lambda} \|\xi_{f,\lambda}\|_2^{p_{u,\lambda}} \leq \delta_\lambda < \infty$ and finite diameter $0 < \nu = \text{diam}(\Xi_\lambda) < \infty$. A finite δ_λ follows from

$$\begin{aligned} \sup_{\xi_{f,\lambda} \in \Xi_\lambda} (\|\xi_{f,\lambda}\|_2^{p_{u,\lambda}})^2 &= \sup_{f \in \mathcal{F}} (\|\xi_{f,\lambda}\|_2^{p_{u,\lambda}})^2 = \sup_{f \in \mathcal{F}} (\|\xi(u_t, f; \lambda)\|_2^{p_{u,\lambda}})^2 = \sup_{f \in \mathcal{F}} \int |\xi(u_t, f; \lambda)|^2 p_{u,\lambda}(u_t) du_t \\ &\leq \sup_{f \in \mathcal{F}} \int |\eta(f; \lambda) \zeta_1(u_t; \lambda) + \zeta_2(u_t; \lambda)|^2 p_{u,\lambda}(u_t) du_t \\ &= \sup_{f \in \mathcal{F}} (\|\eta(f; \lambda) \zeta_1(u_t; \lambda) + \zeta_2(u_t; \lambda)\|_2^{p_{u,\lambda}})^2 \\ &\leq \sup_{f \in \mathcal{F}} (\|\eta(f; \lambda)\| \|\zeta_1(u_t; \lambda)\|_2^{p_{u,\lambda}} + \|\zeta_2(u_t; \lambda)\|_2^{p_{u,\lambda}})^2 \\ &\leq (\|\zeta_1(u_t; \lambda)\|_2^{p_{u,\lambda}})^2 \sup_{f \in \mathcal{F}} \eta(f; \lambda)^2 + (\|\zeta_2(u_t; \lambda)\|_2^{p_{u,\lambda}})^2 \\ &\quad + 2\|\zeta_1(u_t; \lambda)\|_2^{p_{u,\lambda}} \|\zeta_2(u_t; \lambda)\|_2^{p_{u,\lambda}} \sup_{f \in \mathcal{F}} \eta(f; \lambda) < \infty, \end{aligned}$$

where the first inequality holds by the assumption of a separable bound for $|\xi(u_t, f; \lambda)|$, the second by

norm sub-additivity, and the third inequality follows from the assumptions of bounded η and square integrable ζ_1 and ζ_2 (in empirical norm) which ensures that all elements are bounded.

The finite diameter ν is obtained by noting that,

$$\begin{aligned}
\text{diam}(\Xi_\lambda) &= \sup_{(\xi_{f,\lambda}, \xi_{f',\lambda}) \in \Xi_\lambda \times \Xi_\lambda} d_{\Xi_\lambda}(\xi_{f,\lambda}, \xi_{f',\lambda}) \leq \sup_{(\xi_{f,\lambda}, \xi_{f',\lambda}) \in \Xi_\lambda \times \Xi_\lambda} \|\xi_{f,\lambda} - \xi_{f',\lambda}\|_2^{D_{u,\lambda}} \\
&= \sup_{(f, f') \in \mathcal{F} \times \mathcal{F}} \left(\mathbb{E}(\xi_{f,\lambda} - \xi_{f',\lambda})^2 \right)^{\frac{1}{2}} \leq \sup_{(f, f') \in \mathcal{F} \times \mathcal{F}} \left(\mathbb{E}(\eta^*(f, f'; \lambda)^2 \zeta^*(u_t; \lambda)^2) \right)^{\frac{1}{2}} \\
&= \left(\int |\zeta^*(u_t; \lambda)|^2 p_{u,\lambda}(u_t) du_t \right)^{\frac{1}{2}} \sup_{(f, f') \in \mathcal{F} \times \mathcal{F}} \eta^*(f, f'; \lambda) < \infty,
\end{aligned}$$

where the first inequality follows by the assumption that d_{Ξ_λ} is weaker than $\|\cdot\|_2^{D_{u,\lambda}}$, the second inequality follows by the separable bound assumption on the derivative's difference, and the last inequality follows by the assumptions of bounded η^* and empirical norm square integrability of ζ^* . ■