TI 2014-029/III
Tinbergen Institute Discussion Paper

Maximum Likelihood Estimation for Generalized Autoregressive Score Models

Francisco Blasques
Siem Jan Koopman
Andre Lucas

Faculty of Economics and Business Administration, VU University Amsterdam, and Tinbergen Institute, the Netherlands.

Maximum Likelihood Estimation for Generalized Autoregressive Score Models[1]

Francisco Blasques[a], Siem Jan Koopman[a,b], André Lucas[a]

[(a)] *VU University Amsterdam and Tinbergen Institute*

[(b)] *CREATES, Aarhus University*

Abstract

We study the strong consistency and asymptotic normality of the maximum likelihood estimator for a class of time series models driven by the score function of the predictive likelihood. This class of nonlinear dynamic models includes both new and existing observation driven time series models. Examples include models for generalized autoregressive conditional heteroskedasticity, mixed-measurement dynamic factors, serial dependence in heavy-tailed densities, and other time varying parameter processes. We formulate primitive conditions for global identification, invertibility, strong consistency, asymptotic normality under correct specification and under mis-specification. We provide key illustrations of how the theory can be applied to specific dynamic models.

Keywords: time-varying parameter models, GAS, score driven models, Markov processes estimation, stationarity, invertibility, consistency, asymptotic normality.

JEL classifications: C13, C22, C12.

AMS classifications: 62E20 (primary); 62F10, 62F12, 60G10, 62M05, 60H25 (secondary).

1 Introduction

We aim to formulate primitive conditions for global identification, strong consistency and asymptotic normality of the maximum likelihood estimator (MLE)

[1]We thank Peter Boswijk, Christian Francq, Andrew Harvey, and Anders Rahbek, as well as the participants of the "2013 Workshop on Dynamic Models driven by the Score of Predictive Likelihoods", Amsterdam; the "7th International Conference on Computational and Financial Econometrics (CFE 2013)", London; and the "2014 Workshop on Dynamic Models driven by the Score of Predictive Likelihoods", Tenerife, for helpful comments and discussions. Lucas and Blasques thank the Dutch National Science Foundation (NWO; grant VICI453-09-005) for financial support. Email correspondence: f.blasques@vu.nl, s.j.koopman@vu.nl, a.lucas@vu.nl

for the time-invariant parameters in a general class of score driven nonlinear time series models specified by

$$y_t \sim p_y(y_t|f_t;\lambda), \quad f_{t+1} = \omega + \alpha s(f_t, y_t; \lambda) + \beta f_t, \tag{1.1}$$

where y_t is the observed data, f_t is a time varying parameter characterizing the conditional density p_y of y_t, $s(f_t, y_t; \lambda) = S(f_t; \lambda) \cdot \partial \log p_y(y_t|f_t; \lambda)/\partial f_t$ is the scaled score of the predictive conditional likelihood, for some choice of scaling function $S(f_t; \lambda)$, and the static parameters are collected in the vector $\boldsymbol{\theta} = (\omega, \alpha, \beta, \lambda^\top)^\top$ with $^\top$ denoting transposition. This class of models is known as Generalized Autoregressive Score (GAS) models[2] and has been studied by, for example, Creal, Koopman, and Lucas (2011,2013), Harvey (2013), Oh and Patton (2013), Harvey and Luati (2014), Andres (2014), Lucas et al. (2014), and Creal et al. (2014). A well-known special case of (1.1) is the familiar generalized autoregressive conditional heteroskedasticity (GARCH) model of Engle (1982) and Bollerslev (1986),

$$y_t = f_t^{1/2} u_t, \quad f_{t+1} = \omega^* + \alpha^* y_t^2 + \beta^* f_t, \quad u_t \sim \mathrm{N}(0,1), \tag{1.2}$$

where $\{u_t\}$ is a sequence of independently distributed standard normal random variables, and ω^*, α^*, and β^* are static parameters that need to be estimated. Since models (1.1) and (1.2) are both 'observation driven' in the terminology of Cox (1981), the likelihood function is known in closed form through the prediction error decomposition. This facilitates parameter estimation via the method of maximum likelihood (ML).

The choice for y_t^2 in (1.2) to drive changes in f_t, however, is particular to the volatility context. It is not clear what functions of the data one should use in other applications such as the time variation in the shape parameters of a Beta or Gamma distribution. Even for time varying volatility models it is not self-evident that $s(f_t, y_t; \lambda) = y_t^2$ is the best possible choice; see Nelson and Foster (1994) and Creal et al. (2011) for alternative volatility models under fat tails.

The key novelty in equation (1.1) compared to equation (1.2) is the use of the scaled score of the conditional observation density in the updating scheme of the time varying parameter f_t. The modeling framework implied by (1.1) is uniformly applicable whenever a conditional observation density p_y is available. It

[2]Harvey (2013) uses the alternative acronym of Dynamic Conditional Score (DCS) models.

generalizes many familiar dynamic models including nonlinear time series models such as the normal GARCH model, 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), the Beta-t-EGARCH model of Harvey (2013), and many related models. More recently proposed GAS models include the mixed measurement and mixed frequency dynamic factor models of Creal et al. (2014), the multivariate volatility and correlation models for fat-tailed and possibly skewed observations of Creal et al. (2011), Harvey (2013), and Andres (2014), the fat-tailed dynamic (local) level models of Harvey and Luati (2014), and the dynamic copula models of Oh and Patton (2013) and Lucas et al. (2014).

The above references represent a wide range of empirical studies which are based on the GAS model (1.1) and require the maximum likelihood estimation of $\boldsymbol{\theta}$. However, the theoretical properties of the MLE for (1.1) have not been well investigated. This stands in sharp contrast to the large number of results available for the MLE in GARCH models; see, for example, the overviews in Straumann (2005) and Francq and Zakoïan (2010). An additional complexity for the GAS model in comparison to the GARCH model is that the dynamic features of f_t are typically intricate nonlinear functions of lagged y_t's.

We make the following contributions. First, we establish the asymptotic properties of the MLE for GAS models. In particular, we build on the stochastic recurrence equation approach that is used in Bougerol (1993) and Straumann and Mikosch (2006), hereafter referred to as SM06. We obtain the properties of the MLE through an application of the ergodic theorem in Rao (1962) for strictly stationary and ergodic sequences on separable Banach spaces. As in SM06, we use this approach to obtain strong consistency and asymptotic normality of the MLE under mild differentiability requirements and moment conditions. Our results also apply to models outside the class of multiplicative error models (MEM) of Engle (2002) which are considered in SM06. Although our updating equation for the time varying parameter is more specific than the one used in SM06, we present results under more general conditions. For example, the uniform lower bound on the autoregressive updating function adopted in SM06 is only appropriate for the MEM class and is too restrictive in our setting.

Second, we derive the properties of the MLE from primitive *low-level* conditions on the basic structure of the model. Most other contributions in the literature use high-level conditions instead. For example, we do not impose moment conditions on the likelihood function; we obtain the necessary moments from conditions imposed on the updating equation (1.1) directly. Using these weak low-level conditions, we ensure stationarity, ergodicity, invertibility as well as the existence of moments. The use of primitive conditions may be useful for those empirical researchers who want to establish asymptotic properties of the MLE of parameters in their model at hand. The importance of establishing invertibility has been underlined in SM06 and Wintenberger (2013), among others.

Third, we provide primitive *global identification* conditions for the parameters of correctly specified GAS models. In particular we ensure that the likelihood function has a unique maximum over the entire parameter space. Our global results differ from the usual identification results which rely on high-level assumptions and only ensure local identification by relying on the properties of the information matrix at the true parameter value; see, e.g. SM06 and Harvey (2013).

Fourth, all the results above are obtained for large parameter spaces whose boundaries can be derived. Most other consistency and asymptotic normality results typically hold for arbitrarily small parameter spaces containing the true parameter.

Finally, we derive the consistency and asymptotic normality of the MLE for both well-specified and mis-specified GAS models. For the case of mis-specified models, the asymptotic results hold with respect to a pseudo-true parameter that minimizes the Kullback-Leibler divergence between the true unknown probability measure and the measure implied by the model. These results hold despite the potential severity of model mis-specification.

The remainder of our paper is organized as follows. Section 2 introduces the model and establishes notation. In Section 3, we obtain stationarity, ergodicity, invertibility, and existence of moments of filtered GAS sequences from primitive conditions. Section 4 proves global identification, consistency and asymptotic normality of the MLE. In Section 5, we analyze examples using the theory developed in Sections 3 and 4. Section 6 concludes. The proofs of the main theorems are gathered in the Appendix. The proofs of auxiliary propositions and lemmas, together with additional examples, are provided in the Supplementary Appendix (SA).

2 The GAS Model

The generalized autoregressive score model was informally introduced in equation (1.1). For the remainder of the paper, we adopt a more formal description of the model. The GAS model defines the dynamic properties of a d_y-dimensional stochastic sequence $\{y_t\}_{t\in\mathbb{N}}$ given by

$$y_t = g\big(f_t(y^{1:t-1}, \boldsymbol{\theta}, \bar{f})\ ,\ u_t(\lambda)\big), \qquad u_t(\lambda) \sim p_u\big(u_t(\lambda); \lambda\big), \tag{2.1}$$

where $g : \mathcal{F}_g \times \mathcal{U}_g \to \mathcal{Y}_g$ is a link function that is strictly increasing in its second argument, $f_t(y^{1:t-1}, \boldsymbol{\theta}, \bar{f})$ is the time varying parameter function with $y^{1:t-1} = (y_1, \ldots, y_{t-1})$, $\{u_t(\lambda)\}_{t\in\mathbb{N}}$ is an exogenous i.i.d. sequence of random variables for every parameter vector $\lambda \in \Lambda \subseteq \mathbb{R}^{d_\lambda}$, p_u is a density function, and the time varying parameter updating scheme is given by

$$f_{t+1}(y^{1:t}, \boldsymbol{\theta}, \bar{f}) = \omega + \alpha s\big(f_t(y^{1:t-1}, \boldsymbol{\theta}, \bar{f}), y_t\ ;\ \lambda\big) + \beta f_t(y^{1:t-1}, \boldsymbol{\theta}, \bar{f}), \tag{2.2}$$

for $t > 1$, and initialized at $f_1(\emptyset, \boldsymbol{\theta}, \bar{f}) = \bar{f}$, for a nonrandom $\bar{f} \in \mathcal{F} \subseteq \mathbb{R}$, where $\emptyset$ is the empty set, $\boldsymbol{\theta}^\top = (\omega, \alpha, \beta, \lambda^\top) \in \Theta \subseteq \mathbb{R}^{3+d_\lambda}$ is the parameter vector, and $s : \mathcal{F}_s \times \mathcal{Y}_s \times \Lambda \to \mathcal{F}_s$ is the scaled score of the conditional density of y_t given f_t. Whenever possible, we suppress the dependence of $u_t(\lambda)$ and $f_t(y^{1:t-1}, \boldsymbol{\theta}, \bar{f})$ on their arguments, and instead write u_t and f_t, respectively. Also, when there is no risk of confusion, we drop subscripts from the sets $\mathcal{F}_g = \mathcal{F}_s = \mathcal{F}$, so that the functions g and s are assumed to be defined on the support $\mathcal{F}$. We only make a strict separation between these sets when needed, particularly in the proof of our identification result in Theorem 3.

Define $p_y(y_t|f_t; \lambda)$ as the conditional density of y_t given f_t,

$$p_y\big(y_t \mid f_t\ ;\ \lambda\big) = p_u\big(\tilde{g}(f_t, y_t)\ ;\ \lambda\big) \cdot \tilde{g}'(f_t, y_t), \tag{2.3}$$

where $\tilde{g}'_t := \tilde{g}'(f_t, y_t) := \partial\tilde{g}(f_t, y)/\partial y|_{y=y_t}$ is the Jacobian of transformation (2.1) with

$$\tilde{g}_t := \tilde{g}(f_t, y_t) := g^{-1}(f_t, y_t),$$

and $g^{-1}(f_t, y_t)$ denoting the inverse of $g(f_t, u_t)$ with respect to its second argument u_t. The defining aspect of the GAS model is its use of the scaled score function as the driving mechanism in the transition equation (2.2). The scaled score function is defined as

$$s\big(f_t, y_t\ ;\ \lambda\big) = S(f_t; \lambda) \cdot \left[\frac{\partial \tilde{p}_t}{\partial f} + \frac{\partial \tilde{g}'_t}{\partial f}\right], \tag{2.4}$$

with $\tilde{p}_t := \tilde{p}(f_t, y_t; \lambda) = \log p_u(\tilde{g}(f_t, y_t); \lambda)$ and where $S : \mathcal{F}_s \times \Lambda \to \mathcal{F}_s$ is a positive scaling function.

Section 4 establishes existence, consistency and asymptotic normality of the maximum likelihood estimator (MLE) for the vector of parameters $\boldsymbol{\theta}$, where the MLE $\hat{\boldsymbol{\theta}}_T(\bar{f})$ is defined as

$$\hat{\boldsymbol{\theta}}_T(\bar{f}) \in \arg\max_{\boldsymbol{\theta} \in \Theta} \ell_T(\boldsymbol{\theta}, \bar{f}),$$

with the average log likelihood function ℓ_T given by

$$\ell_T(\boldsymbol{\theta}, \bar{f}) = \frac{1}{T} \sum_{t=2}^{T} \Big(\log p_u(\tilde{g}_t; \lambda) + \log \tilde{g}_t' \Big) = \frac{1}{T} \sum_{t=2}^{T} \Big(\tilde{p}_t + \log \tilde{g}_t' \Big). \tag{2.5}$$

The advantage of GAS models is that, similar to other observation driven models, their likelihood function (2.5) is available in closed form and can be computed directly using the GAS measurement and updating equations (2.1) and (2.2), respectively. Consider the following GAS volatility model as an example.

The conditional volatility model

To model the time varying variance of a normal distribution, let p_u be the standard normal density and let $g(f_t, u_t) = f_t^{1/2} u_t$. The score is given by $(y_t^2 - f_t)/(2f_t^2)$. By following Creal et al. (2011, 2013) in scaling the score by the inverse of its conditional expected variance, we obtain $S(f_t; \lambda) = 2f_t^2$. Equation (2.2) reduces to

$$f_{t+1} = \omega + \alpha(y_t^2 - f_t) + \beta f_t. \tag{2.6}$$

Here we recognize the well-known GARCH(1,1) model of Engle (1982) and Bollerslev (1986) as given in equation (1.2), with $\omega^* = \omega$, $\alpha^* = \alpha$, and $\beta^* = \beta - \alpha$. To ensure non-negativity of the variance, we require $\beta > \alpha > 0$ and $\omega > 0$. An alternative for imposing a positive variance is to model the log-variance and to set $g(f_t, u_t) = \exp(f_t/2)u_t$. The inverse conditional variance of the score is then given by $S(f_t; \lambda) = 0.5$. We obtain

$$f_{t+1} = \omega + \alpha \left(\exp(-f_t) y_t^2 - 1 \right) + \beta f_t; \tag{2.7}$$

compare the exponential GARCH (EGARCH) model of Nelson (1991).

The features of the GAS model for volatility can be further illustrated by considering a fat-tailed Student's t density for u_t with zero mean, unit scale, and

$\lambda > 0$ degrees of freedom. Following Creal et al. (2011) for the case $g(f_t, u_t) = f_t^{1/2} u_t$. and scaling the score by the inverse of its conditional variance, $S(f_t; \lambda) = 2(1 + 3\lambda^{-1}) f_t^2$, we obtain

$$f_{t+1} = \omega + \alpha(1 + 3\lambda^{-1}) \left(\frac{(1 + \lambda^{-1}) y_t^2}{1 + \lambda^{-1} y_t^2 / f_t} - f_t \right) + \beta f_t, \tag{2.8}$$

which is the score driven GAS volatility model discussed in Creal et al. (2011, 2013) and Harvey (2013). The model in (2.8) is markedly different from a GARCH model with Student's t innovations, which would still be driven by y_t^2. An advantage of the Student's t conditional score in the GAS transition equation (2.8) is that it mitigates the impact of large values y_t^2 on future values of the variance parameter f_{t+1} through the presence of y_t^2 in the denominator of $s(f_t, y_t; \lambda)$ for $\lambda^{-1} > 0$.

We present further examples of GAS models beyond the volatility context, such as dynamic one-factor models, conditional duration models and time varying regressions models in the Supplementary Appendix.

3 Notation and Preliminary Results

To enable a more convenient exposition, we assume that λ is a scalar, i.e., $d_\lambda = 1$. Given the results in Bougerol (1993) and SM06,[3] we present two related propositions that play their respective roles in the applications of Section 5.

For a scalar random variable x, we define $\|x\|_n := (\mathbb{E}|x|^n)^{1/n}$ for $n > 0$. If the random variable $x(\boldsymbol{\theta})$ depends on a parameter $\boldsymbol{\theta} \in \Theta$, we define $\|x(\cdot)\|_n^\Theta := (\mathbb{E} \sup_{\boldsymbol{\theta} \in \Theta} |x(\boldsymbol{\theta})|^n)^{1/n}$. Furthermore, we define $x^{t_1:t_2} := \{x_t\}_{t=t_1}^{t_2}$, and $x^{t_2} := \{x_t\}_{t=-\infty}^{t_2}$ for any sequence $\{x_t\}_{t\in\mathbb{Z}}$ and any $t_1, t_2 \in \mathbb{N}$. If the sequence $\{x_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ depends on parameter $\boldsymbol{\theta}$, we use short-hand notation $x_{\boldsymbol{\theta}}^{t_1:t_2} := x^{t_1:t_2}(\boldsymbol{\theta})$. Finally, we use $x_t \perp x_t'$ to denote independence between x_t and x_t'.

Propositions 1 and 2 below are written specifically for the GAS model recursion. More general counterparts can be found in the Supplementary Appendix. We first consider the GAS model as driven by u_t rather than y_t to establish results later on for the MLE under a correctly specified GAS model. Define $s_u(f_t, u_t; \lambda) := s(f_t, g(f_t, u_t); \lambda)$ and let $\{f_t(u_\lambda^{1:t-1}, \boldsymbol{\theta}, \bar{f})\}_{t\in\mathbb{N}}$ be generated by

$$f_{t+1}(u_\lambda^{1:t}, \boldsymbol{\theta}, \bar{f}) = \omega + \alpha s_u\big(f_t(u_\lambda^{1:t-1}, \boldsymbol{\theta}, \bar{f}), u_t; \lambda\big) + \beta f_t(u_\lambda^{1:t-1}, \boldsymbol{\theta}, \bar{f}), \tag{3.1}$$

[3] Straumann and Mikosch (2006, Theorem 2.8) extend Bougerol (1993, Theorem 3.1) with the uniqueness of the stationary solution.

for $t > 1$ and initial condition $f_1(\emptyset, \boldsymbol{\theta}, \bar{f}) = \bar{f}$, and with $s_u \in \mathbb{C}^{(1,0,0)}(\mathcal{F}^* \times \mathcal{U} \times \Lambda)$ for some convex $\mathcal{F} \subseteq \mathcal{F}^* \subset \mathbb{R}$. Define the random derivative function $\dot{s}_{u,t}(f^*; \lambda) := \partial s_u(f^*, u_t; \lambda)/\partial f$ and its kth power supremum

$$\rho_t^k(\boldsymbol{\theta}) := \sup_{f^* \in \mathcal{F}^*} |\beta + \alpha\, \dot{s}_{u,t}(f^*; \lambda)|^k.$$

We then have the following proposition.

Proposition 1. *For every $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^{3+d_\lambda}$ let $\{u_t(\lambda)\}_{t\in\mathbb{Z}}$ be an i.i.d. sequence and assume $\exists \bar{f} \in \mathcal{F}$ such that*

(i) $\mathbb{E} \log^+ |s_u(\bar{f}, u_1(\lambda); \lambda)| < \infty$;

(ii) $\mathbb{E} \log \rho_1^1(\boldsymbol{\theta}) < 0$.

Then $\{f_t(u_\lambda^{1:t-1}, \boldsymbol{\theta}, \bar{f})\}_{t\in\mathbb{N}}$ converges e.a.s. to the unique stationary and ergodic (SE) sequence $\{f_t(u_\lambda^{t-1}, \boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ for every $\boldsymbol{\theta} \in \Theta$ as $t \to \infty$.

If furthermore for every $\boldsymbol{\theta} \in \Theta$ $\exists$ $n_f > 0$ such that

(iii) $\|s_u(\bar{f}, u_1(\lambda); \lambda)\|_{n_f} < \infty$;

(iv) $\mathbb{E}\rho_t^{n_f}(\boldsymbol{\theta}) < 1$;

(v) $f_t(u_\lambda^{1:t-1}, \boldsymbol{\theta}, \bar{f}) \perp \rho_t^{n_f}(\boldsymbol{\theta})$ $\forall$ $(t, \bar{f}) \in \mathbb{N} \times \mathcal{F}$.

Then $\sup_t \mathbb{E}|f_t(u_\lambda^{1:t-1}, \boldsymbol{\theta}, \bar{f})|^{n_f} < \infty$ *and* $\mathbb{E}|f_t(u_\lambda^{t-1}, \boldsymbol{\theta})|^{n_f} < \infty$ $\forall$ $\boldsymbol{\theta} \in \Theta$.

Proposition 1 does not only establish stationarity and ergodicity (SE), it also establishes existence of unconditional moments. Conditions *(i)* and *(ii)* in Proposition 1 also provide an almost sure representation of $f_t(u_\lambda^{t-1}, \boldsymbol{\theta})$ in terms of u_λ^{t-1}; see Remark SA.2 in the Supplementary Appendix.

The independence of u_t and $f_t(u_\lambda^{1:t-1}, \boldsymbol{\theta}, \bar{f})$ is sufficient to establish condition *(v)*. We summarize this in Remark 1. The remark also provides a stricter substitute for conditions *(ii)* and *(iv)* based on a straightforward binomial expansion. This stricter condition is often easier to verify for specific models.

Remark 1. If $u_t(\lambda) \perp f_t(u_\lambda^{1:t-1}, \boldsymbol{\theta}, \bar{f})$ $\forall$ $(t, \boldsymbol{\theta}, \bar{f})$, then condition *(v)* in Proposition 1 holds. Furthermore, conditions *(ii)* and *(iv)* can be substituted by the (stricter albeit easier to verify) condition

(iv′) $\sum_{k=0}^{n_f} \binom{n_f}{k} |\alpha|^k \, |\beta|^{n_f - k} \, \mathbb{E} \sup_{f^* \in \mathcal{F}^*} |\dot{s}_{u,t}(f^*; \lambda)|^k < 1$.

Lemma SA.1 and Lemma SA.2 in the Supplemental Appendix present a set of alternative convenient conditions.

Our second proposition is key in establishing moment bounds and e.a.s. convergence of the GAS filtered sequence $\{f_t(y^{1:t-1}, \boldsymbol{\theta}, \bar{f})\}$, uniformly over the parameter space Θ. We prove the result not only for $f_t(y^{1:t-1}, \boldsymbol{\theta}, \bar{f})$ itself, but also for the derivative processes of $f_t(y^{1:t-1}, \boldsymbol{\theta}, \bar{f})$ with respect to $\boldsymbol{\theta}$. These derivative processes play a major role in the proof of asymptotic normality of the MLE later on. Our bounds use only primitive conditions that are formulated directly in terms of the core structure of the model, i.e., in terms of the scaled score s and log density $\tilde{p}$. These primitive conditions use the notion of moment preserving maps, which we define as follows.

Definition 1. (Moment Preserving Maps)
A function $h : \mathbb{R}^q \times \Theta \to \mathbb{R}$ is said to be $\boldsymbol{n}/m$-moment preserving, denoted as $h(\cdot; \boldsymbol{\theta}) \in \mathbb{M}_{\Theta_1,\Theta_2}(\boldsymbol{n}, m)$, if and only if $\mathbb{E} \sup_{\boldsymbol{\theta} \in \Theta_1} |x_{i,t}(\boldsymbol{\theta})|^{n_i} < \infty$ for $\boldsymbol{n} = (n_1, \ldots, n_q)$ and $i = 1, \ldots, q$ implies $\mathbb{E} \sup_{\boldsymbol{\theta} \in \Theta_2} |h(\boldsymbol{x}_t(\boldsymbol{\theta}); \boldsymbol{\theta})|^m < \infty$. If Θ_1 or Θ_2 consists of a singleton, we replace Θ_1 or Θ_2 in the notation by its single element, e.g., $\mathbb{M}_{\boldsymbol{\theta}_1,\Theta_2}$ if $\Theta_1 = \{\boldsymbol{\theta}_1\}$.

For example, every polynomial function $h(x; \boldsymbol{\theta}) = \sum_{j=0}^{J} \theta_j x^j \ \forall \ (x, \boldsymbol{\theta}) \in \mathcal{X} \times \Theta$, $\boldsymbol{\theta} = (\theta_0, \ldots, \theta_J) \in \Theta \subseteq \mathbb{R}^J$ trivially satisfies $h \in \mathbb{M}_{\boldsymbol{\theta},\boldsymbol{\theta}}(n, m)$ with $m = n/J \ \forall \ \boldsymbol{\theta} \in \Theta$. If Θ is compact, then also $h \in \mathbb{M}_{\Theta,\Theta}(n, m)$ with $m = n/J$. Similarly, every k-times continuously differentiable function $h(\cdot; \boldsymbol{\theta}) \in \mathbb{C}^k(\mathcal{X})$ $\forall \ \boldsymbol{\theta} \in \Theta$, with bounded k^{th} derivative $\sup_{x \in \mathcal{X}} |h^{(k)}(x; \boldsymbol{\theta})| \leq \bar{h}_k(\boldsymbol{\theta}) < \infty \ \forall \ \boldsymbol{\theta} \in \Theta$, satisfies $h \in \mathbb{M}_{\boldsymbol{\theta},\boldsymbol{\theta}}(n, m)$ with $m = n/k \ \forall \ \boldsymbol{\theta} \in \Theta$. If furthermore $\sup_{\boldsymbol{\theta} \in \Theta} \bar{h}_k(\boldsymbol{\theta}) \leq \bar{\bar{h}} < \infty$, then $h \in \mathbb{M}_{\Theta,\Theta}(n, m)$ with $m = n/k$; see Lemma SA.6 in the Supplementary Appendix for further details and examples. We note that $\mathbb{M}_{\Theta',\Theta'}(\boldsymbol{n}, m) \subseteq \mathbb{M}_{\Theta,\Theta}(\boldsymbol{n}, m^*)$ for all $m^* \leq m$, and all $\Theta \subseteq \Theta'$.

Moment preservation is a natural requirement in the consistency and asymptotic normality proofs later on, as the likelihood and its derivatives are nonlinear functions of the original data y_t, the time varying parameter $f_t(y^{1:t-1}, \boldsymbol{\theta}, \bar{f})$, and its derivatives with respect to $\boldsymbol{\theta}$.

Consider the GAS recurrence equation in (2.2). Define the random derivative $\dot{s}_{y,t}(f^*; \lambda) := \partial s(f^*, y_t; \lambda)/\partial f$ and the supremum of its kth-power

$$\tilde{\rho}_t^k(\boldsymbol{\theta}) = \sup_{f^* \in \mathcal{F}^*} |\beta + \alpha \, \dot{s}_{y,t}(f^*; \lambda)|^k,$$

with $\mathcal{F} \subseteq \mathcal{F}^* \subset \mathbb{R}$. As mentioned above, the consistency and asymptotic normality proofs also require SE properties of certain derivative processes of $f_t(y^{1:t-1}, \boldsymbol{\theta}, \bar{f})$ with respect to $\boldsymbol{\theta}$. We denote the ith order derivative by $\boldsymbol{f}_t^{(i)}(y^{1:t-1}, \boldsymbol{\theta}, \bar{\boldsymbol{f}}^{0:i})$, which takes values in $\mathcal{F}^{(i)}$, with $\bar{\boldsymbol{f}}^{0:i} \in \mathcal{F}^{(0:i)} = \mathcal{F} \times ... \times \mathcal{F}^{(i)}$ being the fixed initial condition for the first ith order derivatives; see the Supplementary Appendix for further details.

To state Proposition 2 concisely, we write

$$s^{(\boldsymbol{k})}(f, y; \lambda) = \partial^{k_1+k_2+k_3} s(f, y; \lambda)/(\partial f^{k_1} \partial y^{k_2} \partial \lambda^{k_3}),$$

with $\boldsymbol{k} = (k_1, k_2, k_3)$. As $s^{(\boldsymbol{k})}(f, y; \lambda)$ is a function of both the data and the time varying parameter, we impose moment preserving properties on each of the $s^{(\boldsymbol{k})}$, for example, $s^{(\boldsymbol{k})} \in \mathbb{M}_{\Theta,\Theta}(\boldsymbol{n}, n_s^{(\boldsymbol{k})})$, with $n_s^{(\boldsymbol{k})}$ being the number of bounded moments of $s^{(\boldsymbol{k})}$ when its first two arguments have $\boldsymbol{n} := (n_f, n_y)$ moments. We have suppressed the third argument of s, the parameter λ, in the moment preserving properties. We can do so without loss of generality, as λ is not stochastic. We also adopt the more transparent short-hand notation $n_s^f := n_s^{(1,0,0)}$ to denote the preserved moment for the derivative of s with respect to f. Similarly, we define $n_s^{ff} := n_s^{(2,0,0)}$, $n_s^{\lambda} := n_s^{(0,0,1)}$, $n_s^{\lambda\lambda} := n_s^{(0,0,2)}$ and $n_s^{f\lambda} := n_s^{(1,0,1)}$. Using these definitions, we can ensure the existence of the $n_f^{(1)}$th and $n_f^{(2)}$th moments of the first and second derivative of $f_t(y^{1:t-1}, \boldsymbol{\theta}, \bar{f})$ with respect to $\boldsymbol{\theta}$, respectively, where

$$n_f^{(1)} = \min\{n_f, n_s, n_s^{\lambda}\},$$

$$n_f^{(2)} = \min\Big\{n_f^{(1)}\ ,\ n_s^{\lambda}\ ,\ n_s^{\lambda\lambda}\ ,\ \frac{n_s^f n_f^{(1)}}{n_s^f + n_f^{(1)}}\ ,\ \frac{n_s^{ff} n_f^{(1)}}{2n_s^{ff} + n_f^{(1)}}\ ,\ \frac{n_s^{f\lambda} n_f^{(1)}}{n_s^{f\lambda} + n_f^{(1)}}\Big\}.$$

Proposition 2. *Let $\Theta \subset \mathbb{R}^{3+d_\lambda}$ be compact, $s \in \mathbb{C}^{(2,0,2)}(\mathcal{F} \times \mathcal{Y} \times \Lambda)$, and $\{y_t\}_{t\in\mathbb{Z}}$ be an SE sequence satisfying $\mathbb{E}|y_t|^{n_y} < \infty$ for some $n_y \geq 0$. Let $s^{(\boldsymbol{k})} \in \mathbb{M}_{\Theta,\Theta}(\boldsymbol{n}, n_s^{(\boldsymbol{k})})$ with $\boldsymbol{n} := (n_f, n_y)$ such that $n_f^{(1)} > 0$, $n_f^{(2)} > 0$. Finally, assume $\exists\ \bar{\boldsymbol{f}}^{0:i} \in \mathcal{F}^{(0:2)}$ such that*

(i) $\mathbb{E} \log^+ \sup_{\lambda\in\Lambda} |s(\bar{f}, y_t; \lambda)| < \infty$;

(ii) $\mathbb{E} \log \sup_{\boldsymbol{\theta}\in\Theta} \tilde{\rho}_1^1(\boldsymbol{\theta}) < 0$.

Then $\{\boldsymbol{f}_t^{(i)}(y^{1:t-1}, \boldsymbol{\theta}, \bar{\boldsymbol{f}}^{0:i})\}_{t\in\mathbb{N}}$ converges e.a.s. to a unique SE sequence $\{\boldsymbol{f}_t^{(i)}(y^{t-1}, \boldsymbol{\theta})\}_{t\in\mathbb{Z}}$, uniformly on Θ as $t \to \infty$, for $i = 0, 1, 2$.

If furthermore $\exists\ n_f > 0$ such that $n_f^{(1)} > 0$, $n_f^{(2)} > 0$ and

(iii) $\|s(\bar{f}, y_t; \cdot)\|_{n_f}^{\Lambda} < \infty$;

(iv) $\mathbb{E} \sup_{\boldsymbol{\theta}\in\Theta} \tilde{\rho}_1^{n_f}(\boldsymbol{\theta}) < 1$;

(v) $f_t(y^{1:t-1},\boldsymbol{\theta},\bar{f}) \perp \tilde{\rho}_t^{n_f}(\boldsymbol{\theta})\ \forall\ (t,\boldsymbol{\theta},\bar{f})$;

then $\sup_t \|f_t(y^{1:t-1},\cdot,\bar{f})\|_{n_f}^{\Theta} < \infty$, $\|f_t(y^{t-1},\cdot)\|_{n_f}^{\Theta} < \infty$, *and* $\sup_t \|\boldsymbol{f}_t^{(i)}(y^{1:t-1},\cdot,\bar{\boldsymbol{f}}^{0:i})\|_{n_f^{(i)}}^{\Theta} < \infty$ *and* $\|\boldsymbol{f}_t^{(i)}(y^{t-1},\cdot)\|_{n_f^{(i)}}^{\Theta} < \infty$ *for* $i = 1, 2$.

This proposition establishes existence of SE solutions and of unconditional moments for both $\{f_t(y^{1:t-1},\boldsymbol{\theta},\bar{f})\}$ and its first two derivatives. It is useful to make the following observation.

Remark 2. The properties of the sequence $\{f_t(y^{1:t-1},\boldsymbol{\theta},\bar{f})\}$ established in Proposition 2 hold without the assumptions that $s^{(\boldsymbol{k})} \in \mathbb{M}_{\Theta,\Theta}(\boldsymbol{n}, n_s^{(\boldsymbol{k})})$, $n_f^{(1)} > 0$, $n_f^{(2)} > 0$, or $n_f^{(1)} \geq 1$ and $n_f^{(2)} \geq 1$.

The expressions for $n_f^{(1)}$ and $n_f^{(2)}$ appear complex and non-intuitive at first sight. However, they arise naturally from expressions for the derivative processes of $f_t(y^{1:t-1},\boldsymbol{\theta},\bar{f})$ with respect to $\boldsymbol{\theta}$, since they contain sums and products of y_t, $f_t(y^{1:t-1},\boldsymbol{\theta},\bar{f})$, $\boldsymbol{f}_t^{(1)}(y^{1:t-1},\boldsymbol{\theta},\bar{f})$, and transformations thereof. We can simplify the moment requirements substantially by expressing the moments $n_f^{(1)}$ and $n_f^{(2)}$ for the first and second derivative processes in terms of a common minimum moment bound that holds for all derivatives of s. We state this as a separate remark.

Remark 3. Let the assumptions of Proposition 2 hold and define $m_s := \min\{n_s^{(i,0,j)} : (i,j) \in \mathbb{N}_0^2,\ i+j \leq 2\}$. Then the moment bounds on the derivative processes hold with $n_f^{(1)} = m_s$ and $n_f^{(2)} = m_s/3$.

The contraction condition in *(iv)* of Proposition 2 is sometimes difficult to handle. Remark 4 states a set of alternative conditions to bound moments without appealing to *(iv)*; see Proposition SA.2 for a proof.

Remark 4. If $\sup_{(f^*,y;\boldsymbol{\theta})\in\mathcal{F}^*\times\mathcal{Y}\times\Theta} |\beta+\alpha\,\partial s(f^*,y;\lambda)/\partial f| < 1$, we can drop conditions *(iv)* and *(v)* in Proposition 2. Alternatively, *(iv)* and *(v)* in Proposition 2 can be substituted by $\sup_{\boldsymbol{\theta}\in\Theta}\sup_{y\in\mathcal{Y}} |\omega + \alpha s(\bar{f}^*, y; \lambda) + \beta \bar{f}^*| = |\bar{\phi}(\bar{f}^*,\boldsymbol{\theta})| < \infty$ and $\sup_{\boldsymbol{\theta}\in\Theta}\sup_{f^*\in\mathcal{F}^*} |\partial\bar{\phi}(f^*,\boldsymbol{\theta})/\partial f| < 1$, $\mathcal{F} \subseteq \mathcal{F}^*$.

Note that conditions *(iii)* and *(iv)* imply conditions *(i)* and *(ii)*, respectively. Finally, we note that under conditions *(i)* and *(ii)* in Proposition 2, our model is *invertible* as we can write $\boldsymbol{f}_t^{(i)}(y^{t-1},\boldsymbol{\theta})$ as a measurable function of all past observations y^{t-1}; see e.g. Granger and Andersen (1978), SM06 and Wintenberger (2013) and Remark SA.4 in the Supplementary Appendix.

In Section 4 we show that the stochastic recurrence approach followed in Propositions 1 and 2 allows us to obtain consistency and asymptotic normality under weaker differentiability conditions than those typically imposed; see also Section 2.3 of SM06. In particular, instead of relying on the usual pointwise convergence plus stochastic equicontinuity of Andrews (1992) and Potscher and Prucha (1994), we obtain uniform convergence through the application of the ergodic theorem of Rao (1962) for sequences in separable Banach spaces. This constitutes a crucial simplification as working with the third derivative of the likelihood of a general GAS model is typically very cumbersome.

4 Identification, Consistency and Asymptotic Normality

We next formulate the conditions under which the MLE for GAS models is strongly consistent and asymptotically normal. The low-level conditions that we formulate relate directly to the two propositions from Section 3, and particularly to the moment preserving properties. We derive results for both correctly specified and mis-specified models. For a correctly specified model, we are also able to prove a new global identification result from low-level conditions, rather than assuming identification via a high-level assumption.

Assumption 1. $(\Theta, \mathfrak{B}(\Theta))$ *is a measurable space and* Θ *is compact.*

Assumption 2. $\tilde{g} \in \mathbb{C}^{(2,0)}(\mathcal{F} \times \mathcal{Y})$, $\tilde{g}' \in \mathbb{C}^{(2,0)}(\mathcal{F} \times \mathcal{Y})$, $\tilde{p} \in \mathbb{C}^{(2,2)}(\mathcal{G} \times \Lambda)$, *and* $S \in \mathbb{C}^{(2,2)}(\mathcal{F} \times \Lambda)$, *where* $\mathcal{G} := \tilde{g}(\mathcal{Y}, \mathcal{F})$.

The conditions in Assumption 2 are sufficient for $s \in \mathbb{C}^{(2,0,2)}(\mathcal{F} \times \mathcal{Y} \times \Lambda)$. Let Ξ be the event space of the underlying complete probability space. The next theorem establishes the existence of the MLE.

Theorem 1. (Existence) *Let Assumptions 1 and 2 hold. Then there exists a.s. a measurable map* $\hat{\boldsymbol{\theta}}_T(\bar{f}) : \Xi \to \Theta$ *satisfying* $\hat{\boldsymbol{\theta}}_T(\bar{f}) \in \arg\max_{\boldsymbol{\theta} \in \Theta} \ell_T(\boldsymbol{\theta}, \bar{f})$, *for all* $T \in \mathbb{N}$ *and every initialization* $\bar{f} \in \mathcal{F}$.

Let $n_{\log \tilde{g}'}$ and $n_{\tilde{p}}$ define the moment preserving properties of $\log \tilde{g}'$ and $\tilde{p}$, respectively, i.e., let $\log \tilde{g}' \in \mathbb{M}_{\Theta,\Theta}(\boldsymbol{n}, n_{\log \tilde{g}'})$ and $\tilde{p} \in \mathbb{M}_{\Theta,\Theta}(\boldsymbol{n}, n_{\tilde{p}})$ where $\boldsymbol{n} := (n_f, n_y)$. To establish consistency, we use the following two assumptions.

Assumption 3. *$\exists\ \Theta^* \subseteq \mathbb{R}^{3+d_\lambda}$ and $n_f \geq 1$ such that, for every $\bar{f} \in \mathcal{F} \subseteq \mathcal{F}^*$ either*

(i.a) $\|s(\bar{f}, y_t; \cdot)\|_{n_f}^{\Theta^*} < \infty$;

(ii.a) $\sup_{(f^*,y,\boldsymbol{\theta})\in\mathcal{F}^*\times\mathcal{Y}\times\Theta^*} |\beta + \alpha\, \partial s(f^*, y; \lambda)/\partial f| < 1$;

or

(i.b) $\|s(\bar{f}, y_t; \cdot)\|_{n_f}^{\Theta^*} < \infty$;

(ii.b) $\mathbb{E} \sup_{\boldsymbol{\theta}\in\Theta^*} \tilde{\rho}_t^{n_f}(\boldsymbol{\theta}) < 1$;

(iii.b) $f_t(y^{1:t-1}, \boldsymbol{\theta}, \bar{f}) \perp \tilde{\rho}_t^{n_f}(\boldsymbol{\theta})\ \forall\ (t, \boldsymbol{\theta}, \bar{f})$;

or

(i.c) $\sup_{\boldsymbol{\theta}\in\Theta^*} \sup_{y\in\mathcal{Y}} |\omega + \alpha s(\bar{f}, y; \lambda) + \beta \bar{f}^*| = \bar{\phi}(\bar{f}) < \infty$;

(ii.c) $\sup_{f^*\in\mathcal{F}^*} |\partial\bar{\phi}(f^*)/\partial f| < 1$.

Assumption 4. *$n_\ell := \min\{n_{\log \tilde{g}'}, n_{\tilde{p}}\}$ satisfies $n_\ell \geq 1$.*

Assumptions 3 and 4 together ensure the convergence of the sequence $\{f_t(y^{1:t-1}, \boldsymbol{\theta}, \bar{f})\}$ to an SE limit with n_f moments by restricting the moment preserving properties of $\tilde{p}$ and $\log \tilde{g}'$, which determine the core structure of the GAS model. This is achieved through an application of Proposition 2 and Remark 4. Combined with the n_y moments of y_t, we then obtain one bounded moment n_ℓ for the log likelihood function.

Theorem 2. (Consistency) *Let $\{y_t\}_{t\in\mathbb{Z}}$ be an SE sequence satisfying $\mathbb{E}|y_t|^{n_y} < \infty$ for some $n_y \geq 0$ and assume that Assumptions 1-4 hold. Furthermore, let $\boldsymbol{\theta}_0 \in \Theta$ be the unique maximizer of $\ell_\infty(\cdot)$ on the parameter space $\Theta \subseteq \Theta^*$ with Θ^* as introduced in Assumption 3. Then the MLE satisfies $\hat{\boldsymbol{\theta}}_T(\bar{f}) \stackrel{a.s.}{\to} \boldsymbol{\theta}_0$ as $T \to \infty$ for every $\bar{f} \in \mathcal{F}$.*

Theorem 2 shows the strong consistency of the MLE in a mis-specified model setting. Consistency is obtained with respect to a pseudo-true parameter $\boldsymbol{\theta}_0 \in \Theta$ that is *assumed* to be the *unique* maximizer of the limit log likelihood $\ell_\infty(\boldsymbol{\theta})$. This pseudo-true parameter minimizes the Kullback-Leibler divergence between the probability measure of $\{y_t\}_{t\in\mathbb{Z}}$ and the measure implied by the model. The result naturally requires regularity conditions on the observed data $\{y_t\}_{t=1}^T \subset \{y_t\}_{t\in\mathbb{Z}}$ that is generated by an unknown data generating process. Such conditions in this general setting can only be imposed by means of direct assumption. However, under an axiom of correct specification, we can show that y_t has n_y moments and that $\boldsymbol{\theta}_0$ is the unique maximizer of the limit likelihood function. In this case, the properties of the observed data $\{y_t\}_{t=1}^T$ no longer have to be *assumed*. Instead, they can be *derived* from the properties of the GAS model under appropriate restrictions on the parameter space. By establishing 'global identification' we ensure that the limit likelihood has a unique maximum over the entire parameter space rather than only in a small neighborhood of the true parameter. The latter is typically achieved by studying the information matrix.

Define the set $\mathcal{Y}_g \subseteq \mathbb{R}$ as the image of $\mathcal{F}_g$ and U under g; i.e. $\mathcal{Y}_g := \{g(f,u),\ (f,u) \in \mathcal{F}_g \times \mathcal{U}\}$. We recall also that $\mathcal{U}$ denotes the common support of $p_u(\,\cdot\,;\lambda)\ \forall\ \lambda \in \Lambda$ and that $\mathcal{F}_g$, $\mathcal{F}_s$ and $\mathcal{Y}_s$ denote subsets of $\mathbb{R}$ over which the maps g and s are defined, respectively. Below, Λ_* denotes the orthogonal projection of the set $\Theta_* \subseteq \mathbb{R}^{3+d_\lambda}$ onto $\mathbb{R}^{d_\lambda}$. Furthermore, statements for almost every (f.a.e.) element in a set hold with respect to Lebesgue measure. The following two assumptions allow us to derive the appropriate properties for $\{y_t\}_{t\in\mathbb{Z}}$ and to ensure global identification of the true parameter.

Assumption 5. $\exists\ \Theta_* \subseteq \mathbb{R}^{3+d_\lambda}$ *and* $n_u \geq 0$ *such that*

(i) $\mathcal{U}$ *contains an open set for every* $\lambda \in \Lambda_*$;

(ii) $\mathbb{E}\sup_{\lambda\in\Lambda_*} |u_t(\lambda)|^{n_u} < \infty$ *and* $g \in \mathbb{M}(\boldsymbol{n}, n_y)$ *with* $\boldsymbol{n} := (n_f, n_u)$ *and* $n_y \geq 0$.

(iii) $g(f,\cdot) \in \mathbb{C}^1(\mathcal{U})$ *is invertible and* $g^{-1}(f,\cdot) \in \mathbb{C}^1(\mathcal{Y}_g)$ *f.a.e.* $f \in \mathcal{F}_g$;

(iv) $p_y(y|f;\lambda) = p_y(y|f';\lambda')$ *holds f.a.e.* $y \in \mathcal{Y}_g$ *iff* $f = f'$ *and* $\lambda = \lambda'$.

Condition *(i)* of Assumption 5 ensures that the innovations have non-degenerate support. Condition *(ii)* ensures that $y_t(\boldsymbol{\theta}_0)$ has n_y moments when the true f_t has n_f moments. Condition *(iii)* imposes that $g(f,\cdot)$ is continuously differentiable and invertible with continuously differentiable derivative. It ensures that the conditional distribution p_y of y_t given f_t is non-degenerate and uniquely defined by the distribution of u_t. Finally, condition *(iv)* states that the static model defined by the observation equation $y_t = g(f, u_t)$ and the density $p_u(\,\cdot\,;\lambda)$ is identified. It requires the conditional density of y_t given $f_t = f$ to be unique for every pair (f,λ). This requirement is obvious : one would not extend a static model to a dynamic one if the former is not already identified.

Assumption 6. *$\exists\ \Theta_* \subseteq \mathbb{R}^{3+d_\lambda}$ and $n_f > 0$ such that for every $\boldsymbol{\theta} \in \Theta_*$ and every $\bar{f} \in \mathcal{F}_s \subseteq \mathcal{F}_s^*$ either*

(i.a) $\|s_u(\bar{f}, u_1(\lambda); \lambda)\|_{n_f} < \infty$;

(ii.a) $\mathbb{E}\rho_t^{n_f}(\boldsymbol{\theta}) < 1$;

or

(i.b) $\sup_{u\in\mathcal{U}} |s_u(\bar{f}, u; \lambda)| = s_u(\bar{f}; \lambda) < \infty$;

(ii.b) $\sup_{f^*\in\mathcal{F}^*} |\partial s_u(f^*; \lambda)/\partial f| < 1$.

Furthermore, $\alpha \neq 0\ \forall\ \boldsymbol{\theta} \in \Theta$. Finally, for every $(f, \boldsymbol{\theta}) \in \mathcal{F}_s \times \Theta$,

$$\partial s(f, y, \lambda)/\partial y \neq 0, \tag{4.1}$$

for almost every $y \in \mathcal{Y}_g$.

Conditions *(i.a)–(ii.a)* or *(i.b)–(ii.b)* in Assumption 6 ensure that the true sequence $\{f_t(\boldsymbol{\theta}_0)\}$ is SE and has n_f moments by application of Proposition 1 and Remark 1. Together with condition *(iii)* in Assumption 5 we then conclude that the data $\{y_t(\boldsymbol{\theta}_0)\}_{t\in\mathbb{Z}}$ itself is SE and has n_y moments. The inequality stated in (4.1) in Assumption 6, together with the assumption that $\alpha \neq 0$ ensure that the data $\{y_t(\boldsymbol{\theta}_0)\}$ entering the update equation (2.2) renders the filtered $\{f_t\}$ stochastic and non-degenerate.

We can now state the following result.

Theorem 3 (Global Identification)**.** *Let Assumptions 1-6 hold and let the observed data be a subset of the realized path of a stochastic process $\{y_t(\boldsymbol{\theta}_0)\}_{t\in\mathbb{Z}}$ generated by a GAS model under $\boldsymbol{\theta}_0 \in \Theta$. Then $Q_\infty(\boldsymbol{\theta}_0) \equiv \mathbb{E}_{\boldsymbol{\theta}_0}\ell_t(\boldsymbol{\theta}_0) > \mathbb{E}_{\boldsymbol{\theta}_0}\ell_t(\boldsymbol{\theta}) \equiv Q_\infty(\boldsymbol{\theta})\ \forall\ \boldsymbol{\theta} \in \Theta : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$.*

The axiom of correct specification leads us to the global identification result in Theorem 3. We can also use it to establish consistency to the true (rather than pseudo-true) parameter value. This is summarized in the following corollary.

Corollary 1. (Consistency) *Let Assumptions 1-6 hold and $\{y_t\}_{t\in\mathbb{Z}} = \{y_t(\boldsymbol{\theta}_0)\}_{t\in\mathbb{Z}}$ with $\boldsymbol{\theta}_0 \in \Theta$, where $\Theta \subseteq \Theta^* \cap \Theta_*$ with Θ^* and Θ_* defined in Assumptions 3, 5 and 6. Then the MLE $\hat{\boldsymbol{\theta}}_T(\bar{f})$ satisfies $\hat{\boldsymbol{\theta}}_T(\bar{f}) \stackrel{a.s.}{\to} \boldsymbol{\theta}_0$ as $T \to \infty$ for every $\bar{f} \in \mathcal{F}$.*

The consistency region $\Theta^* \cap \Theta_*$ under correct specification is a subset of the consistency region Θ^* for the mis-specified setting. This simply reflects the fact that the axiom of correct specification alone (without parameter space restrictions) is not enough to obtain the desired moment bounds. The parameter space must be restricted as well, to ensure that the GAS model is identified and generates SE data with the appropriate number of moments.

To establish asymptotic normality of the MLE, we make the following assumption.

Assumption 7. *$\exists\ \Theta_*^* \subseteq \mathbb{R}^{3+d_\lambda}$ such that $n_{\ell'} \geq 2$ and $n_{\ell''} \geq 1$, with*

$$n_{\ell'} = \min\left\{ n_{\tilde{p}}^{(0,0,1)} \;,\; \frac{n_{\log \tilde{g}'}^{(1,0)} n_f^{(1)}}{n_{\log \tilde{g}'}^{(1,0)} + n_f^{(1)}} \;,\; \frac{n_{\tilde{p}}^{(1,0,0)} n_f^{(1)}}{n_{\tilde{p}}^{(1,0,0)} + n_f^{(1)}} \right\}, \tag{4.2}$$

$$\begin{aligned} n_{\ell''} = \min\Bigg\{ & n_{\tilde{p}}^{(0,0,2)} \;,\; \frac{n_{\tilde{p}}^{(1,0,1)} n_f^{(1)}}{n_{\tilde{p}}^{(1,0,1)} + n_f^{(1)}} \;,\; \frac{n_{\tilde{p}}^{(2,0,0)} n_f^{(1)}}{2n_{\tilde{p}}^{(2,0,0)} + n_f^{(1)}} \;, \\ & \frac{n_{\tilde{p}}^{(1,0,0)} n_f^{(2)}}{n_{\tilde{p}}^{(1,0,0)} + n_f^{(2)}} \;,\; \frac{n_{\log \tilde{g}'}^{(1,0)} n_f^{(2)}}{n_{\log \tilde{g}'}^{(1,0)} + n_f^{(2)}} \;,\; \frac{n_{\log \tilde{g}'}^{(2,0)} n_f^{(1)}}{2n_{\log \tilde{g}'}^{(2,0)} + n_f^{(1)}} \Bigg\}, \end{aligned} \tag{4.3}$$

$n_f^{(1)}$ and $n_f^{(2)}$ as defined above Proposition 2, $s^{(\boldsymbol{k})} \in \mathbb{M}_{\Theta_*^*,\Theta_*^*}(\boldsymbol{n}, n_s^{(\boldsymbol{k})})$, $\tilde{p}^{(\boldsymbol{k}')} \in \mathbb{M}_{\Theta_*^*,\Theta_*^*}(n_{\tilde{g}}, n_{\tilde{p}}^{(\boldsymbol{k}')})$, $(\log \tilde{g}')^{(\boldsymbol{k}'')} \in \mathbb{M}_{\Theta_*^*,\Theta_*^*}(\boldsymbol{n}, n_{\log \tilde{g}'}^{(\boldsymbol{k}'')})$, and $\boldsymbol{n} := (n_f, n_y)$,

Similar to Proposition 2, the moment conditions in Assumption 7 might seem cumbersome at first. The expressions follow directly, however, from the expressions for the derivatives of the log likelihood with respect to $\boldsymbol{\theta}$. Consider the expression for $n_{\ell'}$ in (4.2) as an example. The first term in the derivative of $\ell_T(\boldsymbol{\theta}, \bar{f})$ with respect to $\boldsymbol{\theta}$ is the derivative of the log-density with respect to the static parameter λ. Its moments are ensured by $n_{\tilde{p}}^{(0,0,1)}$. The second term is the derivative of the log Jacobian with respect to f_t, multiplied (via the chain rule) by the derivative of f_t with respect to λ. Moment preservation is ensured by the second term in (4.2) involving $n_{\log \tilde{g}'}^{(0,1)}$ and $n_f^{(1)}$ through the application of a standard Hölder inequality. The same reasoning applies to the third component which corresponds to the derivative of $\tilde{p}_t$ with respect to f_t, multiplied by the derivative of f_t with respect to λ. The expressions in Assumption 7 can be simplified considerably to a single moment condition as stated in the following remark.

Remark 5. Let m denote the lowest of the primitive derivative moment numbers $n_{\tilde{p}}^{(1,0,0)}$, $n_{\tilde{p}}^{(1,0,1)}$, $n_{\log \tilde{g}'}^{(1,0)}$, etc. Then $m \geq 4$ implies $n_{\ell'} \geq 2$ and $n_{\ell''} \geq 1$.

It is often just as easy, however, to check the moment conditions formulated in Assumption 7 directly rather than the simplified conditions in Remark 5; see Section 5.

The following theorem states the main result for asymptotic normality of the MLE under mis-specification, with $\text{int}(\Theta)$ denoting the interior of Θ.

Theorem 4. (Asymptotic Normality) *Let $\{y_t\}_{t\in\mathbb{Z}}$ be an SE sequence satisfying $\mathbb{E}|y_t|^{n_y} < \infty$ for some $n_y \geq 0$ and let Assumptions 1–4 and 7 hold. Furthermore, let $\boldsymbol{\theta}_0 \in \text{int}(\Theta)$ be the unique maximizer of $\ell_\infty(\boldsymbol{\theta})$ on Θ, where $\Theta \subseteq \Theta^* \cap \Theta^*_*$ with Θ^* and Θ^*_* as defined in Assumptions 3 and 7. Then, for every $\bar{f} \in \mathcal{F}$, the ML estimator $\hat{\boldsymbol{\theta}}_T(\bar{f})$ satisfies*

$$\sqrt{T}(\hat{\boldsymbol{\theta}}_T(\bar{f}) - \boldsymbol{\theta}_0) \xrightarrow{d} \mathrm{N}\big(0, \mathcal{I}^{-1}(\boldsymbol{\theta}_0)\mathcal{J}(\boldsymbol{\theta}_0)\mathcal{I}^{-1}(\boldsymbol{\theta}_0)\big) \quad as \quad T \to \infty,$$

where $\mathcal{I}(\boldsymbol{\theta}_0) := \mathbb{E}\tilde{\ell}''_t(\boldsymbol{\theta}_0)$ is the Fisher information matrix, $\tilde{\ell}_t(\boldsymbol{\theta}_0)$ denotes the log likelihood contribution of the tth observation evaluated at $\boldsymbol{\theta}_0$, and $\mathcal{J}(\boldsymbol{\theta}_0) := \mathbb{E}\tilde{\ell}'_t(\boldsymbol{\theta}_0)\tilde{\ell}'_t(\boldsymbol{\theta}_0)^\top$ is the expected outer product of gradients.

For a correctly specified model, we have the following corollary.

Corollary 2. (Asymptotic Normality) *Let Assumptions 1-7 hold and assume $\{y_t(\boldsymbol{\theta}_0)\}_{t\in\mathbb{Z}}$ is a random sequence generated by a GAS model under some $\boldsymbol{\theta}_0 \in \text{int}(\Theta)$ where $\Theta \subseteq \Theta^* \cap \Theta_* \cap \Theta^*_*$ with Θ^*, Θ_* and Θ^*_* defined in Assumptions 3 and 5-7. Then, for every $\bar{f} \in \mathcal{F}$, the MLE $\hat{\boldsymbol{\theta}}_T(\bar{f})$ satisfies*

$$\sqrt{T}(\hat{\boldsymbol{\theta}}_T(\bar{f}) - \boldsymbol{\theta}_0) \xrightarrow{d} \mathrm{N}\big(0, \mathcal{I}^{-1}(\boldsymbol{\theta}_0)\big) \quad as \quad T \to \infty,$$

with $\mathcal{I}(\boldsymbol{\theta}_0)$ the Fisher information matrix defined in Theorem 4.

We next apply the results to a range of different GAS models.

5 Applications of GAS ML Theory

The illustrations below show how the theory of Section 4 can be applied to real models. In particular, we show how the theory is applied to models with different observation equations, innovation densities and time varying parameters f_t with nonlinear dynamics. Due to space considerations, additional examples are presented in the Supplemental Appendix; see Blasques et al. (2014b).

5.1 Time Varying Mean for the Skewed Normal

The GAS location model $y_t = f_t + u_t$ has been studied extensively by Harvey (2013) and Harvey and Luati (2014). We consider an example where u_t is drawn from the skewed normal distribution with unit scale, see O'Hagan and Leonard (1976). For a multivariate GAS volatility example using skewed distributions, we refer to Lucas et al. (2014). We have $p_u(u_t;\lambda) = 2p_{\mathrm{N}}(u_t)P_{\mathrm{N}}(\lambda\, u_t)$, with p_{N} and P_{N} denoting the standard normal pdf and cdf, respectively, and $\lambda \in [-1,1]$ denoting the skewness parameter. We use the scaling function $S(f_t;\lambda) \equiv 1$. In this case, the GAS recursion is given by (2.2) with

$$s(f_t, y_t; \lambda) = (y_t - f_t) \cdot \left(1 - \alpha^2 \, \frac{p_{\mathrm{N}}\big(\lambda\,(y_t - f_t)\big)^2}{P_{\mathrm{N}}\big(\lambda\,(y_t - f_t)\big)}\right). \tag{5.1}$$

For $\lambda = 0$, the score collapses to the residual $y_t - f_t$, which is the natural driver for the mean of a symmetric normal distribution. For $\lambda \neq 0$, the GAS update is nonlinear in f_t. For example, for $\lambda > 0$, the skewed normal distribution is right skewed and the score assigns less importance to positive $y_t - f_t$. This is very intuitive: for $\lambda > 0$, we expect to see relatively more cases of $y_t > f_t$ versus $y_t < f_t$. Therefore, observation $y_t > f_t$ should not have a strong impact on the update for f_t conmpared to observation $y_t < f_t$. The converse holds for $\lambda < 0$. This is similar to the asymmetry in the GAS dynamics obtained for the generalized hyperbolic skewed t distribution in the volatility case; see Lucas et al. (2014).

5.1.1 Local Results Under Correct Specification

When we assume that the model is correctly specified, we can replace $(y - f_t)$ in (5.1) by u_t. We directly obtain that $s_u(f_t, u_t; \lambda)$ is independent of f_t, and therefore $\dot{s}_u(f_t, u_t; \lambda) = 0$ and $\rho_t^k(\boldsymbol{\theta}) = |\beta|$ for all k. All other conditions are easily verified. For any point $\boldsymbol{\theta}_0$ inside the region $|\beta| < 1$, we thus obtain local consistency and asymptotic normality in a small ball around $\boldsymbol{\theta}_0$; compare Harvey and Luati (2014).

5.1.2 Global Results Under Correct Specification

We can establish model invertibility and regions for global identification, consistency and asymptotic normality for the MLE by using the theory from Section 4. Since

$$\tilde{\rho}_t^k(\boldsymbol{\theta}) \approx \max\left\{|\beta - \alpha\,(1 - 0.436\lambda^2)|\;,\; |\beta - \alpha\,(1 + 0.289\lambda^2)|\right\}^k, \tag{5.2}$$

is independent of y_t (see the Supplemental Appendix for details), model invertibility, the asymptotic SE results and the existence of moments of f_t, but also of its derivatives, are ensured as long as $\tilde{\rho}_t^1(\boldsymbol{\theta}) < 1$. Given (5.1), we can set $n_{\log \tilde{g}'}$ arbitrary large and $n_{\tilde{p}} = \min(n_y, n_f)/2$, such that we require $n_y \geq 2$ for consistency. This is ensured if both $|\beta| < 1$ and (5.2) hold. As both conditions are independent of y_t, we also obtain asymptotic normality in the same region. Global identification also follows since Assumptions 5 and 6 hold trivially.

5.1.3 Global Results Under Mis-Specification

By Theorems 2 and 4, under mis-specification, we can drop the requirement $|\beta| < 1$ and just retain condition 5.2 under the assumption that y_t is SE and has unconditional second moments.

5.2 Fat-tailed duration models with logarithmic link function

Models for intertemporally correlated arrival times were initiated by Engle and Russell (1998) using the Weibull based autoregressive conditional duration (ACD) model and extended to the Burr distrubution by Grammig and Maurer (2000). Bauwens and Giot (2000) study a logarithmic version of the ACD model. Consider a duration model $y_t = \exp(f_t)u_t$ with fat-tailed distribution

$$p_u(u_t) = \left(1 + \lambda^{-1}u_t\right)^{-\lambda-1}, \tag{5.3}$$

such that $\mathbb{E}[u_t] = 1 - \lambda^{-1}$ if $\lambda > 1$. A potential drawback of the exponential link function is that the contraction properties are not always easy to verify; compare the discussion of the EGARCH case in SM06.

To simplify the resulting expressions, we scale[4] the score by $(1 + \lambda^{-1})^{-1}$. The scaled score function for the GAS update equation (2.2) and its random derivative are then given by

$$s(f_t, y_t; \lambda) = \frac{e^{-f_t}y_t}{1 + \lambda^{-1}e^{-f_t}y_t} - 1, \quad \dot{s}_{y,t}(f_t; \lambda) = \frac{-e^{-f_t}y_t}{(1 + \lambda^{-1}e^{-f_t}y_t)^2}, \tag{5.4}$$

respectively. It further implies that $s_u(f_t, u_t; \lambda) = u_t/(1 + \lambda^{-1}u_t) - 1$ and $\dot{s}_{u,t}(f_t; \lambda) = 0$. We can use these expressions directly to check the properties of the MLE.

[4]We can also scale by the inverse conditional variance of the score, $1 + 2\lambda^{-1}$, without affecting the main result, but making the resulting expressions more cumbersome.

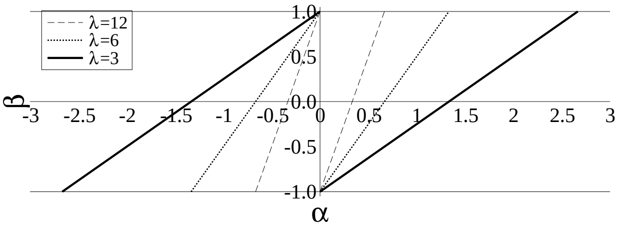

Figure 1: Local and global consistency regions for fat-tailed duration model for different λ.

5.2.1 Local Results Under Correct Specification

Since $\dot{s}_{u,t}(f_t;\lambda)=0$, it follows immediately that $\rho_t^k(\boldsymbol{\theta})=|\beta|$; see also Blasques et al. (2012) and Harvey (2013). The moment preserving properties are checked easily. As a result, for any $\boldsymbol{\theta}_0$ such that $|\beta|<1$ in Figure 1 we obtain that the MLE is consistent and asymptotically normal in a small neighborhood of $\boldsymbol{\theta}_0$. This makes the model under correct specification markedly different from the EGARCH case; see also SM06.

5.2.2 Global Results Under Correct Specification

Consider first the case of an exponential distribution ($\lambda\to\infty$). Using (5.4), $\tilde{\rho}_t^k(\boldsymbol{\theta})$ collapses to $\sup_{f\in\mathcal{F}}|\beta-\alpha\exp(-f)y_t|$, which is unbounded for fixed y_t if $\alpha>0$, unless we impose a lower bound on f_t. The latter can be done by imposing $\omega\geq\underline{\omega}\in\mathbb{R}$ and $\beta>\alpha>0$ and picking an appropriate starting value $\bar{f}$. These parameter restrictions result in a non-degenerate SE region and therefore they are often imposed in practice for the EGARCH model; compare with SM06.

For $\lambda<\infty$, we need not impose such restrictions. In this case Assumptions 5 and 6 are easily satisfied and global identification is obtained directly. Next, we use (5.4) and obtain that $\tilde{\rho}_t^k(\boldsymbol{\theta})=\max(|\beta|,|\beta-\alpha\lambda/4|)$. This maximum is obtained for $f=\log(y_t/\lambda)$ and is independent of y_t itself. Due to this independence, the same parameter restrictions apply for model invertibility and SE as well as for the existence of moments n_f of any order. To obtain global consistency and asymptotic normality, we therefore need $n_f\geq 1$ and $n_f\geq 2$ for $n_\ell,n_{\ell''}\geq 1$ and $n_{\ell'}\geq 2$, respectively. The region where $\tilde{\rho}_t^2(\boldsymbol{\theta})<1$ are plotted in Figure 1 for several values of λ.

5.2.3 Results Under Mis-Specification

By Theorem 2 and 4, as the supremum $\tilde{\rho}_t^2(\boldsymbol{\theta})$ does not depend on y_t, the regions for consistency and asymptotic normality are identical under correct and incorrect specification.

5.3 Gaussian Time Varying Conditional Volatility Models

When considering a normal distribution with time varying variance f_t, the GAS model, with scale equals the inverse of its conditional variance, coincides with the GARCH model (1.2). Stationarity, consistency, and asymptotic normality conditions for GARCH models have been well studied in the literature; see, for example, the original contributions of Lee and Hansen (1994) and Lumsdaine (1996), and references in the extensive reviews provided by Straumann (2005) and Francq and Zakoïan (2010). The GARCH model is based on $\tilde{p}_t = -0.5 \log f_t - 0.5 y_t^2 / f_t$ and can be expressed as

$$y_t = g(f_t, u_t) = h(f_t) u_t = f_t^{1/2} u_t, \quad u_t \sim p_u(u_t; \lambda). \tag{5.5}$$

5.3.1 Local Results Under Correct Specification

When model (5.5) is correctly specified, we have the stochastic recurrence equation (3.1) with $s_u(f_t, u_t; \lambda) = (u_t^2 - 1) f_t$. Part *(ii)* of Proposition 1 implies that y_t is SE if $\mathbb{E} \log \rho_1^1(\boldsymbol{\theta}) < 0$. Since $\rho_1^1(\boldsymbol{\theta}) = |(\beta - \alpha) + \alpha u_t^2|$, it reduces to the familiar Nelson (1990) condition $\mathbb{E} \log |\beta^* + \alpha^* u_t^2| < 0$, with $\beta^* = \beta - \alpha$ and $\alpha^* = \alpha$. In this same region, we can ensure that $n_f > 0$. Consistency then follows as the likelihood function under correct specification is logarithmic in f_t and quadratic in u_t.

We note that $n_f \geq 1$ holds if $\mathbb{E} \rho_t^1(\boldsymbol{\theta}) = \mathbb{E}|(\beta - \alpha) + \alpha u_t^2| = |\beta| < 1$. This produces the familiar triangle $0 < \beta = \beta^* + \alpha^* < 1$. Furthermore $n_f \geq 2$ holds if $\mathbb{E} \rho_t^2(\boldsymbol{\theta}) = \mathbb{E}|(\beta - \alpha) + \alpha u_t^2|^2 = \beta^2 + 2\alpha^2 < 1$. We thus recover all standard local consistency and asymptotic normality results; see Blasques et al. (2014a) for further details.

5.3.2 Global Results Under Correct Specification

Next we show how to establish invertibility of the model, (global) identification, strong consistency and asymptotic normality results outside a small neighborhood of $\boldsymbol{\theta}_0$. For strong consistency, we verify Assumptions 3 and 4. As

$s(\bar{f}, y_t; \lambda) = y_t^2 - \bar{f}$, we obtain $\tilde{\rho}_t^{n_f}(\boldsymbol{\theta}) = |\beta - \alpha|$ for arbitrary n_f, such that Assumption 3*(i)* holds as long as $|\beta - \alpha| < 1$ and $n_f \leq n_y/2$. Let $\omega \geq \underline{\omega} > 0$, such that f_t is uniformly bounded from below for an appropriate initialization $\bar{f} > 0$. If $n_y \geq 2$, Assumption 4 also holds with $n_{\log \tilde{g}'}$ arbitrarily large, $n_{\tilde{p}} = n_y/2$, and $n_\ell = n_y/2 \geq 1$. As shown above, under correct specification $n_y = 2$ if $n_f = 1$, i.e., in the entire triangle $1 > \beta > \alpha > 0$. For any Θ that is a compact subset of this triangle, the MLE is globally strongly consistent. The model is also globally identified for points inside this triangle area since Assumptions 5 and 6 hold. For asymptotic normality, we require $n_{\ell'} \geq 2$ in Assumption 7. We can set $n_{\tilde{p}}^{(0,0,1)}$, $n_{\log \tilde{g}'}^{(1,0)}$, and n_s^λ arbitrarily large, while $n_f = n_s = n_f^{(1)} = n_{\tilde{p}}^{(1,0,0)} = n_y/2$. As a result, we obtain $n_{\ell'} = n_y/4$, such that $n_{\ell'} \geq 2$ requires $n_y \geq 8$. Under correct specification, $n_y \geq 8$ requires $n_f \geq 4$. The latter exists using proposition 1 if $\mathbb{E}\rho_t^4(\boldsymbol{\theta}) < 1$, which is ensured for every (α, β) on the set $\{(\alpha, \beta) \mid \beta > \alpha > 0 \text{ and } \beta^4 + 12\alpha^2\beta^2 + 32\alpha^3\beta + 60\alpha^4 < 1\}$. For any Θ that is a compact subset of this region, the MLE is (globally) asymptotically normally distributed.

5.3.3 Results under Mis-Specification

Theorems 2 and 4 imply that, under incorrect specification, the MLE is globally strongly consistent for any compact subset inside the region $1 + \alpha > \beta > \alpha > 0$ as long as we assume that the data is SE with $n_y \geq 2$. We obtain global asymptotic normality over the same region if $n_y \geq 8$.

5.4 Student's t Time Varying Conditional Volatility Models

Let $\{u_t\}_{t \in \mathbb{N}}$ be fat-tailed by assuming that $u_t \sim t(0, 1; \lambda)$ for the model $y_t = h(f_t)u_t$. If $h(f_t) = \exp(f_t/2)$, parameter updates for a correctly specified model become linear in f_t. Harvey (2013) explored local asymptotic properties of the MLE for this model. As in Creal at al. (2011,2013) and Lucas et al. (2014), we consider the model $y_t = f_t^{1/2} u_t$, with its scaling equals the inverse information. The GAS update of volatility is given by (2.2) with $s(f_t, y_t; \lambda) = \left(1 + 3\lambda^{-1}\right)\left(\frac{(1+\lambda^{-1})y_t^2}{1+y_t^2/(\lambda f_t)} - f_t\right)$; see the Supplemental Appendix for further details. The asymptotic properties for the MLE in the above model have not been investigated before.

5.4.1 Local Results Under Correct Specification

For a correctly specified model, we obtain $\rho_t^k(\boldsymbol{\theta}) = (\beta + \alpha \dot{s}_{u,t}(f_t;\lambda))^k$, where the absolute values have been dropped because $\beta > (1+3\lambda^{-1})\alpha > 0$ and $\dot{s}_{u,t}(f_t;\lambda) \geq -(1+3\lambda^{-1})$ for all u_t, and the supremum has been dropped because $\dot{s}_{u,t}(f_t;\lambda)$ does not depend on f_t. Note that $\lambda^{-1}u_t^2/(1+\lambda^{-1}u_t^2)$ is Beta$(1/2, \lambda/2)$ distributed, such that we can express the moments of $\rho_t^k(\boldsymbol{\theta})$ in analytical form; see also Harvey (2013). For the first and second moment of f_t (and its derivatives) to exist we require $\mathbb{E}\big|\beta + \alpha \dot{s}_{u,t}(f_t;\lambda)\big| = \mathbb{E}\big[\beta + \alpha \dot{s}_{u,t}(f_t;\lambda)\big] = \beta < 1$ and $\mathbb{E}|\beta + \alpha \dot{s}_{u,t}(f_t;\lambda)|^2 = \beta^2 + 2\alpha^2(1+\lambda^{-1})^2(1+3\lambda^{-1}) < 1$. For every $\boldsymbol{\theta}$ in a small neighborhood of $\boldsymbol{\theta}_0$ satisfying the contraction condition, we can establish the local identification, consistency, and asymptotic normality of the MLE. Note that these regions apply even if $\lambda > 0$ is arbitrarily small. In this case, hardly any moments of the data exist, yet still $n_f^{(i)} \geq 2$, $i = 0, 1, 2$.[5] This makes the current model substantially different from the Student's t GARCH model. For the latter the second and fourth order moment of u_t would need to exist to ensure the first and second order moment of f_t.

5.4.2 Global Results Under Correct Specification

Due to the uniform boundedness of $s(f_t, y_t; \lambda)$ in y_t, Assumption 3 is satisfied for arbitrary n_f. Moreover, we have $\tilde{\rho}_t^k(\boldsymbol{\theta}) \leq \big(\beta + \alpha\,(\lambda+3)\big)^k$, for any t and k due to the uniform boundedness of $\dot{s}_{y,t}(f_t;\lambda)$ (see Supplemental Appendix) in both y_t and f_t. Assumption 4 holds with $n_{\tilde{g}} = n_y$. Due to the logarithmic form of $\tilde{p}$ and $\log \tilde{g}'$ in f_t and y_t, we can set $n_{\log \tilde{g}'}$ and $n_{\tilde{p}}$ arbitrarily large as long as $n_f > 0$ and $n_y > 0$, respectively. The existence and global consistency of the MLE follow immediately if $\beta + \alpha\,(\lambda + 3) < 1$. Global asymptotic normality in addition requires $(\beta + \alpha\,(\lambda+3))^2 < 1$ due to Assumption 7. For identification, we notice that the Assumptions 5 and 6 are again satisfied by the same argument as for the normal GAS volatility model.

5.4.3 Global Results Under Mis-Specification

Though easy to operate, the uniform bound $(\beta + \alpha(\lambda+3))^2 < 1$ may imply only a small global consistency and asymptotic normality region for the MLE, particularly if λ is allowed to be large. The uniform boundedness of $\tilde{\rho}_t^k(\boldsymbol{\theta})$, however,

[5] As shown in Proposition SA.1, this is due to the boundedness of the score function of the Student's t distribution that drives the volatility dynamics in the correctly specified case.

implies that the expectation in the contraction condition $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\tilde{\rho}_t^k(\boldsymbol{\theta}) < 1$ can be consistently estimated by the sample average. Such estimated regions are typically substantially larger than the region implied by the uniform bound. More details as well as examples of estimated regions for global asymptotic normality for the Student's t GAS volatility model can be found in the Supplemental Appendix.

6 Conclusions

In this paper we have developed an asymptotic theory for the properties of the maximum likelihood estimator (MLE) in a new class of score driven models that we refer to as generalized autoregressive score (GAS) models. The GAS model has recently been proposed and successfully applied in a range of empirical analyses. The current paper complements the earlier applied literature on GAS models by formally proving the asymptotic properties of the MLE for such models, such as global identification, consistency, and asymptotic normality. The asymptotic properties were provided for both well-specified and mis-specified model settings. Our theorems use primitive, low-level conditions that refer directly to the functions that make up the core of the GAS model. We also stated conditions under which the GAS model is invertible. For the case of correctly specified models, we were able to establish a global identification result outside a small neighborhood containing the true parameter. We believe that our results establish the proper foundation for ML estimation and hypothesis testing for the GAS model in empirical work.

A Proofs of Theorems

Proof of Theorem 1. Assumption 2 implies that $\ell_T(\boldsymbol{\theta}, \bar{f})$ is a.s. continuous (a.s.c.) in $\boldsymbol{\theta} \in \Theta$ through continuity of each $\tilde{\ell}_t(\boldsymbol{\theta}, \bar{f}) = \ell(f_t, y, \boldsymbol{\theta})$, ensured in turn by the differentiability of $\tilde{p}$, $\tilde{g}$, $\tilde{g}'$, the implied a.s.c. of $s(f_t, y; \lambda) = \partial\tilde{p}_t/\partial f$ in $(f_t; \lambda)$ and the resulting continuity of f_t in $\boldsymbol{\theta}$ as a composition of t continuous maps. The compactness of Θ implies by Weierstrass' theorem that the arg max set is non-empty a.s. and hence that $\hat{\boldsymbol{\theta}}_T$ exists a.s. $\forall\ T \in \mathbb{N}$. Similarly, Assumption 2 implies that $\ell_T(\boldsymbol{\theta}, \bar{f}) = \ell\big(\{y_t\}_{t=1}^T, \{f_t\}_{t=1}^T, \boldsymbol{\theta}\big)$ continuous in y_t $\forall\ \boldsymbol{\theta} \in \Theta$ and hence measurable w.r.t. a Borel σ-algebra. The measurability of $\hat{\boldsymbol{\theta}}_T$ follows from White (1994, Theorem 2.11) or Gallant and White (1988, Lemma 2.1, Theorem

2.2). $\square$

Proof of Theorem 2. Following the classical consistency argument that is found e.g. in White (1994, Theorem 3.4) or Gallant and White (1988, Theorem 3.3), we obtain $\hat{\boldsymbol{\theta}}_T(\bar{f}) \overset{a.s.}{\to} \boldsymbol{\theta}_0$ from the uniform convergence of the criterion function and the identifiable uniqueness of the maximizer $\boldsymbol{\theta}_0 \in \Theta$

$$\sup_{\boldsymbol{\theta}:\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|>\epsilon} \ell_\infty(\boldsymbol{\theta}) < \ell_\infty(\boldsymbol{\theta}_0) \ \forall \ \epsilon > 0.$$

Step 1, uniform convergence: Let $\ell_T(\boldsymbol{\theta})$ denote the likelihood function $\ell_T(\boldsymbol{\theta}, \bar{f})$ with $f_t(y^{1:t-1}, \boldsymbol{\theta}, \bar{f})$ replaced by $f_t(y^{t-1}, \boldsymbol{\theta})$. Also define $\ell_\infty(\boldsymbol{\theta}) = \mathbb{E}\tilde{\ell}_t(\boldsymbol{\theta}) \ \forall \ \boldsymbol{\theta} \in \Theta$, with $\tilde{\ell}_t$ denoting the contribution of the tth observation to the likelihood function ℓ_T We have

$$\begin{aligned} \sup\nolimits_{\boldsymbol{\theta}\in\Theta} |\ell_T(\boldsymbol{\theta}, \bar{f}) - \ell_\infty(\boldsymbol{\theta})| \ &\leq \\ \sup\nolimits_{\boldsymbol{\theta}\in\Theta} |\ell_T(\boldsymbol{\theta}, \bar{f}) - \ell_T(\boldsymbol{\theta})| + \sup\nolimits_{\boldsymbol{\theta}\in\Theta} |\ell_T(\boldsymbol{\theta}) - \ell_\infty(\boldsymbol{\theta})|&. \end{aligned} \quad \text{(A.1)}$$

The first term vanishes by the convergence of $f_t(y^{1:t-1}, \boldsymbol{\theta}, \bar{f})$ to $f_t(y^{t-1}, \boldsymbol{\theta})$ and a continuous mapping argument, and the second by Rao (1962).

For the first term in (A.1), we show that $\sup_{\boldsymbol{\theta}\in\Theta} |\tilde{\ell}_t(\boldsymbol{\theta}, \bar{f}) - \tilde{\ell}_t(\boldsymbol{\theta})| \overset{a.s.}{\to} 0$ as $t \to \infty$. The expression for the likelihood in (2.5) and the differentiability conditions in Assumption 2 ensure that $\tilde{\ell}_t(\cdot, \bar{f}) = \ell(f_t(y^{1:t-1}, \cdot, \bar{f}), y_t, \cdot)$ is continuous in $(f_t(y^{1:t-1}, \cdot, \bar{f}), y_t)$. Using Remark 2, all the assumptions of Proposition 2 relevant for the process $\{f_t\}$ hold as well. To see this, note that the compactness of Θ is imposed in Assumption 1; the moment bound $\mathbb{E}|y_t|^{n_y} < \infty$ is ensured in the statement of Theorem 2; the differentiability $s \in \mathbb{C}^{(2,0,2)}(\mathcal{F} \times \mathcal{Y} \times \Lambda)$ is implied by $\tilde{g} \in \mathbb{C}^{(2,0)}(\mathcal{F} \times \mathcal{Y})$, $\tilde{p} \in \mathbb{C}^{(2,2)}(\mathcal{G} \times \Lambda)$, and $S \in \mathbb{C}^{(2,2)}(\mathcal{F} \times \Lambda)$); and finally, conditions *(i)*-*(v)* in Proposition 2 are ensured by Assumption 3. Note that under the alternative set of conditions proposed in Assumption 3, we can use Remark 4 and drop conditions *(iv)* *(v)* in Proposition 2. As a result, there exists a unique SE sequence $\{f_t(y^{1:t-1}, \cdot)\}_{t\in\mathbb{Z}}$ such that $\sup_{\boldsymbol{\theta}\in\Theta} |f_t(y^{1:t-1}, \boldsymbol{\theta}, \bar{f}) - f_t(y^{t-1}, \boldsymbol{\theta})| \overset{a.s.}{\to} 0 \ \forall \bar{f} \in \mathcal{F}$, and $\sup_t \mathbb{E} \sup_{\boldsymbol{\theta}\in\Theta} |f_t(y^{1:t-1}, \boldsymbol{\theta}, \bar{f})|^{n_f} < \infty$ and $\mathbb{E} \sup_{\boldsymbol{\theta}\in\Theta} |f_t(y^{t-1}, \boldsymbol{\theta})|^{n_f} < \infty$ with $n_f \geq 1$. Hence, the first term in (A.1) strongly converges to zero by an application of the continuous mapping theorem for $\ell : \mathbb{C}(\Theta, \mathcal{F}) \times \mathcal{Y} \times \Theta \to \mathbb{R}$.

For the second term in (A.1), we apply the ergodic theorem for separable Banach spaces of Rao (1962) (see also Straumann and Mikosch (2006, Theorem 2.7)) to the sequence $\{\ell_T(\cdot)\}$ with elements taking values in $\mathbb{C}(\Theta)$, so that $\sup_{\boldsymbol{\theta}\in\Theta} |\ell_T(\boldsymbol{\theta}) - \ell_\infty(\boldsymbol{\theta})| \overset{a.s.}{\to} 0$ where $\ell_\infty(\boldsymbol{\theta}) = \mathbb{E}\tilde{\ell}_t(\boldsymbol{\theta}) \ \forall \ \boldsymbol{\theta} \in \Theta$. The ULLN $\sup_{\boldsymbol{\theta}\in\Theta} |\ell_T(\boldsymbol{\theta}) - \mathbb{E}\tilde{\ell}_t(\boldsymbol{\theta})| \overset{a.s.}{\to} 0$ as $T \to \infty$ follows, under a moment bound $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |\tilde{\ell}_t(\boldsymbol{\theta})| < \infty$, by the SE nature of $\{\ell_T\}_{t\in\mathbb{Z}}$, which is implied by continuity of ℓ on the SE sequence $\{(f_t(y^{t-1}, \cdot), y_t)\}_{t\in\mathbb{Z}}$ and Proposition 4.3 in Krengel (1985). The moment bound $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |\tilde{\ell}_t(\boldsymbol{\theta})| < \infty$ is ensured by $\sup_{\boldsymbol{\theta}\in\Theta} \mathbb{E}|f_t(y^{t-1}, \boldsymbol{\theta})|^{n_f} < \infty \ \forall \ \boldsymbol{\theta} \in \Theta$, $\mathbb{E}|y_t|^{n_y} < \infty$, and the fact that Assumption 3 implies $\ell \in \mathbb{M}(\boldsymbol{n}, n_\ell)$ with $\boldsymbol{n} = (n_f, n_y)$ and $n_\ell \geq 1$.

Step 2, uniqueness: Identifiable uniqueness of $\boldsymbol{\theta}_0 \in \Theta$ follows from for example White (1994) by the assumed uniqueness, the compactness of Θ, and the continuity of the limit $\mathbb{E}\tilde{\ell}_t(\boldsymbol{\theta})$ in $\boldsymbol{\theta} \in \Theta$, which is implied by the continuity of ℓ_T in $\boldsymbol{\theta} \in \Theta \ \forall \ T \in \mathbb{N}$ and the uniform convergence of the objective function proved earlier. □

Proof of Theorem 3. We index the true $\{f_t\}$ and the observed random sequence $\{y_t\}$ by the parameter $\boldsymbol{\theta}_0$, e.g. $\{y_t(\boldsymbol{\theta}_0)\}$, since under the correct specification assumption the observed data is a subset of the realized path of a stochastic process $\{y_t\}_{t\in\mathbb{Z}}$ generated by a GAS model under $\boldsymbol{\theta}_0 \in \Theta$. First note that by Proposition 1 the true sequence $\{f_t(\boldsymbol{\theta}_0)\}$ is SE and has at least n_f moments for any $\boldsymbol{\theta} \in \Theta$. Conditions *(i)* and *(ii)* of Proposition 1 hold immediately by Assumption 6 and condition *(v)* follows immediately from the i.i.d. exogenous nature of the sequence $\{u_t\}$. The SE nature and n_f moments of $\{f_t(\boldsymbol{\theta}_0)\}$ together with part *(iii)* of Assumption 5 imply, in turn, that $\{y_t(\boldsymbol{\theta}_0)\}$ is SE with n_y moments.

Step 1, formulation and existence of the limit criterion $Q_\infty(\boldsymbol{\theta})$: As shown in the proof of Theorem 2, the limit criterion function $Q_\infty(\boldsymbol{\theta})$ is now well-defined for every $\boldsymbol{\theta} \in \Theta$ by

$$Q_\infty(\boldsymbol{\theta}) = \mathbb{E}\tilde{\ell}_t(\boldsymbol{\theta}) = \mathbb{E}\log p_{y_t|y^{t-1}}\Big(y_t(\boldsymbol{\theta}_0)\Big|y^{t-1}(\boldsymbol{\theta}_0); \boldsymbol{\theta}\Big).$$

As a normalization, we subtract the constant $Q_\infty(\boldsymbol{\theta}_0)$ from $Q_\infty(\boldsymbol{\theta})$ and focus on showing that

$$Q_\infty(\boldsymbol{\theta}) - Q_\infty(\boldsymbol{\theta}_0) < 0 \ \forall \ (\boldsymbol{\theta}_0, \boldsymbol{\theta}) \in \Theta \times \Theta : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0.$$

Using the dynamic structure of the GAS model, we can substitute the conditioning on $y^{t-1}(\boldsymbol{\theta}_0)$ above by a conditioning on $f_t(y^{t-1}(\boldsymbol{\theta}_0); \boldsymbol{\theta})$, with the random variable $f_t(y^{t-1}(\boldsymbol{\theta}_0); \boldsymbol{\theta})$ taking values in $\mathcal{F}$ through the recursion

$$f_{t+1}\big(y^t(\boldsymbol{\theta}_0); \boldsymbol{\theta}\big) = \phi\Big(f_t\big(y^{t-1}(\boldsymbol{\theta}_0); \boldsymbol{\theta}\big), y_t(\boldsymbol{\theta}_0); \boldsymbol{\theta}\Big) \ \forall \ t \in \mathbb{Z}.$$

Under the present conditions, the limit process $\big\{f_t(y^{t-1}(\boldsymbol{\theta}_0);\boldsymbol{\theta})\big\}_{t\in\mathbb{Z}}$ is a measurable function of $y^{t-1}(\boldsymbol{\theta}_0) = \{y_{t-1}(\boldsymbol{\theta}_0), y_{t-2}(\boldsymbol{\theta}_0), \ldots,\}$, and hence SE by Krengel's theorem for any $\boldsymbol{\theta} \in \Theta$; see also SM06.[6] For the sake of this proof, we adopt the shorter notation

$$\tilde{f}_t(\boldsymbol{\theta}_0, \boldsymbol{\theta}) \equiv f_t\big(y^{t-1}(\boldsymbol{\theta}_0), \boldsymbol{\theta}\big), \qquad f_t(\boldsymbol{\theta}_0, \boldsymbol{\theta}_0) \equiv \tilde{f}_t(\boldsymbol{\theta}_0, \boldsymbol{\theta}_0),$$

and substitute the conditioning on $y^{t-1}(\boldsymbol{\theta}_0)$ by a conditioning on $f_t(\boldsymbol{\theta}_0, \boldsymbol{\theta}_0)$ and $f_t(\boldsymbol{\theta}_0, \boldsymbol{\theta})$. We obtain

$$\begin{aligned} Q_\infty(\boldsymbol{\theta}) - Q_\infty(\boldsymbol{\theta}_0) &= \mathbb{E} \log p_{y_t|f_t}\Big(y_t(\boldsymbol{\theta}_0)\Big| f_t(\boldsymbol{\theta}_0, \boldsymbol{\theta}); \lambda\Big) \\ &\qquad - \mathbb{E} \log p_{y_t|f_t}\Big(y_t(\boldsymbol{\theta}_0)\Big| f_t(\boldsymbol{\theta}_0, \boldsymbol{\theta}_0); \lambda_0\Big) \\ &= \int\!\!\int\!\!\int \log \frac{p_{y_t|f_t}(y|\tilde{f};\lambda)}{p_{y_t|f_t}(y|f;\lambda_0)} \, \mathrm{d}P_{y_t,f_t,\tilde{f}_t}(y, f, \tilde{f}; \boldsymbol{\theta}_0, \boldsymbol{\theta}), \end{aligned} \tag{A.2}$$

$\forall\ (\boldsymbol{\theta}_0, \boldsymbol{\theta}) \in \Theta \times \Theta : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, with $P_{y_t,f_t,\tilde{f}_t}(y, f, \tilde{f}; \boldsymbol{\theta}_0, \boldsymbol{\theta})$ denoting the cdf of $(y_t(\boldsymbol{\theta}_0), f_t(\boldsymbol{\theta}_0, \boldsymbol{\theta}_0), \tilde{f}_t(\boldsymbol{\theta}_0, \boldsymbol{\theta}))$. Define the bivariate cdf $P_{f_t,\tilde{f}_t}(f, \tilde{f}; \boldsymbol{\theta}_0, \boldsymbol{\theta})$ for the pair $(f_t(\boldsymbol{\theta}_0, \boldsymbol{\theta}_0), \tilde{f}_t(\boldsymbol{\theta}_0, \boldsymbol{\theta}))$. Note that the cdf $P_{f_t,\tilde{f}_t}(f, \tilde{f}; \boldsymbol{\theta}_0, \boldsymbol{\theta})$ depends on $\boldsymbol{\theta}$ through the recursion defining $\tilde{f}_t(\boldsymbol{\theta}_0, \boldsymbol{\theta})$, and on $\boldsymbol{\theta}_0$ through $y^t(\boldsymbol{\theta}_0)$ and $f_t(\boldsymbol{\theta}_0, \boldsymbol{\theta}_0)$. Also note that for any $(\boldsymbol{\theta}_0, \boldsymbol{\theta}) \in \Theta \times \Theta$ this cdf does not depend on the initialization $\bar{f}_1$ because, under the present conditions, the limit criterion is a function of the unique limit SE process $\big\{f_t(y^{t-1}(\boldsymbol{\theta}_0);\boldsymbol{\theta})\big\}_{t\in\mathbb{Z}}$, and not of $\big\{f_t(y^{1:t-1}(\boldsymbol{\theta}_0);\boldsymbol{\theta}, \bar{f}_1)\big\}_{t\in\mathbb{N}}$, which depends on $\bar{f}_1$; see the proof of Theorem 2.

We re-write the normalized limit criterion function $Q_\infty(\boldsymbol{\theta}) - Q_\infty(\boldsymbol{\theta}_0)$ by factorizing the joint distribution $P_{y_t,f_t,\tilde{f}_t}(y, f, \tilde{f}, \boldsymbol{\theta}_0, \boldsymbol{\theta})$ as

$$\begin{aligned} P_{y_t,f_t,\tilde{f}_t}(y, f, \tilde{f}; \boldsymbol{\theta}_0, \boldsymbol{\theta}) &= P_{y_t|f_t,\tilde{f}_t}(y|f, \tilde{f}; \boldsymbol{\theta}_0, \boldsymbol{\theta}) \cdot P_{f_t,\tilde{f}_t}(f, \tilde{f}; \boldsymbol{\theta}_0, \boldsymbol{\theta}) \\ &= P_{y_t|f_t}(y|f, \lambda_0) \cdot P_{f_t,\tilde{f}_t}(f, \tilde{f}; \boldsymbol{\theta}_0, \boldsymbol{\theta}), \end{aligned}$$

where the second equality holds because under the axiom of correct specification, and conditional on $f_t(\boldsymbol{\theta}_0, \boldsymbol{\theta}_0)$, observed data $y_t(\boldsymbol{\theta}_0)$ does not depend on $f_t(\boldsymbol{\theta}_0, \boldsymbol{\theta})\ \forall\ (\boldsymbol{\theta}_0, \boldsymbol{\theta}) \in \Theta \times \Theta : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$. We also note that the conditional distribution $P_{y_t|f_t}(y|f, \lambda_0)$ has a density $p_{y_t|f_t}(y|f, \lambda_0)$ defined in equation (2.3). The existence of this density follows because $g(f, \cdot)$ is a diffeomorphism $g(f, \cdot) \in \mathbb{D}(\mathcal{U})$ for every $f \in \mathcal{F}$, i.e., it is continuously differentiable and uniformly invertible

[6] $f_t(\cdot;\boldsymbol{\theta})$ is a measurable map from $\mathcal{Y}^{t-1}$ to $\mathcal{F}$ where where $\mathcal{Y}^{t-1} = \prod_{\tau\in\mathbb{Z}:\tau\leq t} \mathcal{Y}$ and its measure maps elements of $\mathfrak{B}(\mathcal{Y}^{t-1})$ to the interval $[0, 1]\ \forall\ \boldsymbol{\theta} \in \Theta$. The random variable $f_t(y^{t-1}(\boldsymbol{\theta}_0);\boldsymbol{\theta})$, on the other hand, maps elements of $\mathfrak{B}(\mathcal{F})$ to the interval $[0, 1]$.

with differentiable inverse.[7]

We can now re-write $Q_\infty(\boldsymbol{\theta}) - Q_\infty(\boldsymbol{\theta}_0)$ as

$$
\begin{aligned}
&Q_\infty(\boldsymbol{\theta}) - Q_\infty(\boldsymbol{\theta}_0) = \\
&\int\int\int \log \frac{p_{y_t|f_t}(y|\tilde{f};\lambda)}{p_{y_t|f_t}(y|f;\lambda_0)} \, \mathrm{d}P_{y_t|f_t}(y|f,\lambda_0) \cdot \mathrm{d}P_{f_t,\tilde{f}_t}(f,\tilde{f};\boldsymbol{\theta}_0,\boldsymbol{\theta}) = \\
&\int\int\left[\int \log \frac{p_{y_t|f_t}(y|\tilde{f};\lambda)}{p_{y_t|f_t}(y|f;\lambda_0)} \, \mathrm{d}P_{y_t|f_t}(y|f,\lambda_0)\right] \mathrm{d}P_{f_t,\tilde{f}_t}(f,\tilde{f};\boldsymbol{\theta}_0,\boldsymbol{\theta}) = \\
&\int\int\left[\int p_{y_t|f_t}(y|f,\lambda_0) \log \frac{p_{y_t|f_t}(y|\tilde{f};\lambda)}{p_{y_t|f_t}(y|f;\lambda_0)} \, \mathrm{d}y\right] \mathrm{d}P_{f_t,\tilde{f}_t}(f,\tilde{f};\boldsymbol{\theta}_0,\boldsymbol{\theta}),
\end{aligned}
$$

$\forall\ (\boldsymbol{\theta}_0,\boldsymbol{\theta}) \in \Theta \times \Theta : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$.

Step 2, use of Gibb's inequality: The Gibbs inequality ensures that, for any given $(f,\tilde{f},\lambda_0,\lambda) \in \mathcal{F} \times \mathcal{F} \times \Lambda \times \Lambda$, the inner integral above satisfies

$$
\int p_{y_t|f_t}(y|f,\lambda_0) \log \frac{p_{y_t|f_t}(y|\tilde{f};\lambda)}{p_{y_t|f_t}(y|f;\lambda_0)} \, \mathrm{d}y \leq 0,
$$

with strict equality holding if and only if $p_{y_t|f_t}(y|\tilde{f};\lambda) = p_{y_t|f_t}(y|f;\lambda_0)$ almost everywhere in $\mathcal{Y}$ w.r.t. $p_{y_t|f_t}(y|f,\lambda_0)$. As such, the strict inequality $Q_\infty(\boldsymbol{\theta}) - Q_\infty(\boldsymbol{\theta}_0) < 0$ holds if and only if, for every pair $(\boldsymbol{\theta}_0,\boldsymbol{\theta}) \in \Theta \times \Theta$, there exists a set $\mathit{YFF} \subseteq \mathcal{Y} \times \mathcal{F} \times \mathcal{F}$ containing triplets $(y,f,\tilde{f})$ with $f \neq \tilde{f}$ and with orthogonal projections $\mathit{YF} \subseteq \mathcal{Y} \times \mathcal{F}$ and $\mathit{FF} \subseteq \mathcal{F} \times \mathcal{F}$, etc., satisfying

(i) $p_{y_t|f_t}(y|f,\lambda_0) > 0\ \forall\ (y,f) \in \mathit{YF}$;

(ii) if $(\tilde{f},\lambda) \neq (f,\lambda_0)$, then $p_{y_t|f_t}(y|\tilde{f};\lambda) \neq p_{y_t|f_t}(y|f;\lambda_0)\ \forall\ (y,f,\tilde{f}) \in \mathit{YFF}$;

(iii) if $\lambda = \lambda_0$ and $(\omega,\alpha,\beta) \neq (\omega_0,\alpha_0,\beta_0)$, then $P_{f_t,\tilde{f}_t}(f,\tilde{f};\boldsymbol{\theta}_0,\boldsymbol{\theta}) > 0$ for every $(f,\tilde{f}) \in \mathit{FF}\ :\ \tilde{f} \neq f$.

Step 2A, check conditions (i) and (ii): Condition (i) follows by noting that under the correct specification axiom, the conditional density $p_{y_t|f_t}(y|f,\lambda_0)$ is implicitly defined by $y_t(\boldsymbol{\theta}_0) = g(f,u_t)$, $u_t \sim p_u(u_t;\lambda_0)$. Note that $g(f,\cdot)$ is a diffeomorphism $g(f,\cdot) \in \mathbb{D}(\mathcal{U})$ for every $f \in \mathcal{F}_g$ and hence an open map, i.e., $g^{-1}(f,U) \in \mathcal{T}(\mathcal{Y}_g)$ for every $U \in \mathcal{T}(\mathcal{Y}_g)$ where $\mathcal{T}(\mathbb{A})$ denotes a topology on the set $\mathbb{A}$. Therefore, since $p_u(u;\lambda) > 0\ \forall\ (u,\lambda) \in \mathcal{U} \times \Lambda$ with $\mathcal{U}$ containing an

[7] The same however cannot be said of the distribution $P_{f_t,\tilde{f}_t}(f,\tilde{f};\boldsymbol{\theta}_0,\boldsymbol{\theta})$. Even though the sequence $\{f_t(\boldsymbol{\theta}_0,\boldsymbol{\theta},f_1)\}_{t\in\mathbb{N}}$ admits a density for every $(\boldsymbol{\theta}_0,\boldsymbol{\theta}) \in \Theta \times \Theta$, the limit sequence $\{f_t(\boldsymbol{\theta}_0,\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ may fail to posses one.

open set by assumption, we obtain that $\exists Y \in \mathcal{T}(\mathcal{Y}_g)$ such that $p_{y_t|f_t}(y|f,\lambda_0) > 0 \;\forall\; (y,f) \in Y \times \mathcal{F}_g$, namely the image of any open set $U \subseteq \mathcal{U}$ under $g(f,\cdot)$.

Condition (ii) is implied directly by the assumption that $p_{y|f_t}(y|f,\lambda) = p_{y|f_t}(y|f',\lambda')$ almost everywhere in $\mathcal{Y}$ if and only if $f = f' \wedge \lambda = \lambda'$. Note that we use condition (ii) to impose $\lambda = \lambda_0$ in condition (iii), as we already have $Q_\infty(\boldsymbol{\theta}_0) > Q_\infty(\boldsymbol{\theta})$ for any $\boldsymbol{\theta} \in \Theta$ such that $\lambda \neq \lambda_0$, regardless of whether $\tilde{f} \neq f$ or $\tilde{f} = f$.

Step 2B, check condition (iii): Before attempting to prove condition (iii), we note that if condition (i) holds, then the set F cannot be a singleton. This follows from the fact that under condition (i) the set Y must contain an open set. Since $\alpha \neq 0 \;\forall\; \boldsymbol{\theta} \in \Theta$, and since for every $(f,\lambda) \in \mathcal{F} \times \Lambda$ we have $\partial s(f,y,\lambda)/\partial y \neq 0$ almost everywhere in Y_s, we conclude that s is an open map. As a result, conditional on $\tilde{f}_t(\boldsymbol{\theta}_0,\boldsymbol{\theta}) = f$, we have that $\tilde{f}_{t+1}(\boldsymbol{\theta}_0,\boldsymbol{\theta})$ is a continuous random variable with density $p_{\tilde{f}_{t+1}|\tilde{f}_t}(\boldsymbol{\theta}_0,\boldsymbol{\theta})$ that is strictly positive on some open set F^* (i.e. the image of Y under ϕ). Furthermore, since this holds for every $f_t(\boldsymbol{\theta}_0,\boldsymbol{\theta}) = f$, it also holds regardless of the marginal of $f_t(\boldsymbol{\theta}_0,\boldsymbol{\theta})$. This implies that F is not a singleton.

Condition (iii) is obtained by a proof by contradiction. In particular, we note that, for every pair $(\boldsymbol{\theta}_0,\boldsymbol{\theta}) \in \Theta \times \Theta : \lambda = \lambda_0 \wedge (\omega,\alpha,\beta) \neq (\omega_0,\alpha_0,\beta_0)$, if there exists no set $F\!F \subseteq \mathcal{F} \times \mathcal{F}$ satisfying $f \neq \tilde{f} \;\forall\; (f,\tilde{f}) \in F\!F$ such that $P_{f_t,\tilde{f}_t}(f,\tilde{f},\boldsymbol{\theta}_0,\boldsymbol{\theta}) > 0 \;\forall\; (f,\tilde{f}) \in F\!F$, then it must be that $(\omega,\alpha,\beta) = (\omega_0,\alpha_0,\beta_0)$. The proof goes as follows. Let $(\boldsymbol{\theta}_0,\boldsymbol{\theta}) \in \Theta \times \Theta$ be a pair satisfying $\lambda = \lambda_0 \wedge (\omega,\alpha,\beta) \neq (\omega_0,\alpha_0,\beta_0)$. If there exists no set $F\!F \subseteq \mathcal{F} \times \mathcal{F}$ that is an orthogonal projection of $Y\!F\!F$ and satisfies $f \neq \tilde{f}$ and $P_{f_t,\tilde{f}_t}(f,\tilde{f};\boldsymbol{\theta}_0,\boldsymbol{\theta}) > 0 \;\forall\; (f,\tilde{f}) \in F\!F$, then for almost every event $e \in \mathcal{E}$ there exists a point $f_e \in \mathcal{F}$ such that $\tilde{f}_t(\boldsymbol{\theta}_0,\boldsymbol{\theta}) \overset{a.s.}{=} f_t(\boldsymbol{\theta}_0,\boldsymbol{\theta}_0) = f_e$ and $\tilde{f}_{t+1}(\boldsymbol{\theta}_0,\boldsymbol{\theta}) \overset{a.s.}{=} f_{t+1}(\boldsymbol{\theta}_0,\boldsymbol{\theta}_0)$ for any $t \in \mathbb{Z}$ of our choice. This, in turn, implies that for every $(\boldsymbol{\theta}_0,\boldsymbol{\theta}) \in \Theta \times \Theta : \lambda = \lambda_0 \wedge (\omega,\alpha,\beta) \neq (\omega_0,\alpha_0,\beta_0)$ we have

$$
\begin{aligned}
\phi(f_e,y_e,\boldsymbol{\theta}) - \phi(f_e,y_e,\boldsymbol{\theta}_0) &= (\omega-\omega_0) + (\beta-\beta_0)f_e + (\alpha-\alpha_0)s(f_e,y_e,\lambda_0) \\
&= (\omega-\omega_0) + (\beta-\beta_0)f_e \\
&\qquad + (\alpha-\alpha_0)\Big(s(f_e,y^*,\lambda_0) + \frac{\partial s(f_e,y_e^{**},\lambda_0)}{\partial y}(y_e-y^*)\Big) \\
&= A_0 + A_1(y_e)(y_e-y^*) = 0,
\end{aligned}
$$

with

$$A_0 := (\omega - \omega_0) + (\beta - \beta_0) f_e + (\alpha - \alpha_0) s(f_e, y^*, \lambda_0),$$
$$A_1(y_e) := (\alpha - \alpha_0) \frac{\partial s(f_e, y_e^{**}, \lambda_0)}{\partial y},$$

where we used the mean value theorem,

$$s(f_e, y_e, \lambda_0) = s(f_e, y^*, \lambda_0) + \frac{s(f_e, y_e^{**}, \lambda_0)}{\partial y}(y_e - y^*),$$

and with A_1 a function of y_e (through $y^{**} = y^{**}(y_e)$). Note that A_0 does not depend on y_e. The condition $A_0 + A_1(y_e)(y_e - y_e^*) = 0 \;\forall\; y_e \in Y$ holds if and only if $A_0 = 0$ and $A_1(y_e) = 0 \;\forall\; y_e \in Y$. Note that the case where the update is not a function of y_e because $A_1(y_e) = (y_e - y_e^*)^{-1}$ is ruled out by assumption by the fact that $\alpha \neq 0 \;\forall\; \boldsymbol{\theta} \in \Theta$ and that $\partial s(f, y, \lambda)/\partial y \neq 0$ for every $\lambda \in \Lambda$ and almost every $(y, f) \in \mathcal{Y}_s \times \mathcal{F}_s$. As a result, $A_1(y_e) = 0 \;\forall\; y_e \in Y$ if and only if $\alpha = \alpha_0$.

Finally, given $\alpha = \alpha_0 \wedge \lambda = \lambda_0$, the condition that $A_0 = 0$ now reduces to $A_0 := (\omega - \omega_0) + (\beta - \beta_0) f_e$. Hence, by the same argument, we have that $A_0 = 0 \;\Leftrightarrow\; (\omega_0 - \omega) + (\beta_0 - \beta) f_e = 0$ can only hold for every f_e on a non-singleton set F if and only if $\omega = \omega_0$ and $\beta = \beta_0$. This establishes the desired contradiction and hence we conclude that condition (iii) must hold. As a result, an open set $YFF \subseteq \mathcal{Y} \times \mathcal{F} \times \mathcal{F}$ with properties (i)–(iii) exists, and therefore $Q_\infty(\boldsymbol{\theta}) - Q_\infty(\boldsymbol{\theta}_0) < 0$ holds with strict inequality for every pair $(\boldsymbol{\theta}_0, \boldsymbol{\theta}) \in \Theta \times \Theta$. □

Proof of Corollary 1. The desired result is obtained by showing (i) that under the maintained assumptions, $\{y_t\}_{t\in\mathbb{Z}} \equiv \{y_t(\boldsymbol{\theta}_0)\}_{t\in\mathbb{Z}}$ is an SE sequence satisfying $\mathbb{E}|y_t(\boldsymbol{\theta}_0)|^{n_y} < \infty$; (ii) that $\boldsymbol{\theta}_0 \in \Theta$ is the unique maximizer of $\ell_\infty(\boldsymbol{\theta}, \bar{f})$ on Θ; and then (iii) appealing to Theorem 2. The fact that $\{y_t(\boldsymbol{\theta}_0)\}_{t\in\mathbb{Z}}$ is an SE sequence is obtained by applying Proposition 1 under Assumptions 5 and 6 to ensure that $\{f_t(y^{1:t-1}, \boldsymbol{\theta}_0, \bar{f})\}_{t\in\mathbb{N}}$ converges e.a.s. to an SE limit $\{f_t(y^{1:t-1}, \boldsymbol{\theta}_0\}_{t\in\mathbb{Z}}$ satisfying $\mathbb{E}|f_t(y^{1:t-1}, \boldsymbol{\theta}_0)|^{n_f} < \infty$. This implies by continuity of g on $\mathcal{F} \times \mathcal{U}$ (implied by $\tilde{g} \in \mathbb{C}^{(2,0)}(\bar{\mathcal{F}} \times \mathcal{Y})$ in Assumption 2) that $\{y_t(\boldsymbol{\theta}_0)\}_{t\in\mathbb{Z}}$ is SE. Furthermore, $g \in \mathbb{M}_{\boldsymbol{\theta},\boldsymbol{\theta}}(\boldsymbol{n}^*, n_y)$ with $\boldsymbol{n}^* = (n_f, n_u)$ in Assumption 5 implies that $\mathbb{E}|y_t(\boldsymbol{\theta}_0)|^{n_y} < \infty$. Finally, the uniqueness of $\boldsymbol{\theta}_0$ is obtained by applying Theorem 3 under Assumptions 5 and 6. □

Proof of Theorem 4. Following the classical proof of asymptotic normality found e.g. in White (1994, Theorem 6.2)), we obtain the desired result from: (i) the strong consistency of $\hat{\boldsymbol{\theta}}_T \xrightarrow{a.s.} \boldsymbol{\theta}_0 \in \text{int}(\Theta)$; (ii) the a.s. twice continuous differentiability of $\ell_T(\boldsymbol{\theta}, \bar{f})$ in $\boldsymbol{\theta} \in \Theta$; (iii) the asymptotic normality of the score

$$\sqrt{T}\ell'_T\big(\boldsymbol{\theta}_0, \boldsymbol{f}_1^{(0:1)}\big) \xrightarrow{d} \text{N}\big(0, \mathcal{J}(\boldsymbol{\theta}_0)\big), \qquad \mathcal{J}(\boldsymbol{\theta}_0) = \mathbb{E}\big(\tilde{\ell}'_t(\boldsymbol{\theta}_0)\tilde{\ell}'_t(\boldsymbol{\theta}_0)^\top\big); \tag{A.3}$$

(iv) the uniform convergence of the likelihood's second derivative,

$$\sup_{\boldsymbol{\theta}\in\Theta} \big\|\ell''_T(\boldsymbol{\theta}, \boldsymbol{f}_1^{(0:2)}) - \ell''_\infty(\boldsymbol{\theta})\big\| \xrightarrow{a.s.} 0; \tag{A.4}$$

and finally, (v) the non-singularity of the limit $\ell''_\infty(\boldsymbol{\theta}) = \mathbb{E}\tilde{\ell}''_t(\boldsymbol{\theta}) = \mathcal{I}(\boldsymbol{\theta})$.

Step 1, consistency and differentiability: The consistency condition $\hat{\boldsymbol{\theta}}_T \xrightarrow{a.s.} \boldsymbol{\theta}_0 \in \text{int}(\Theta)$ in (i) follows under the maintained assumptions by Theorem 2 and the additional assumption that $\boldsymbol{\theta}_0 \in \text{int}(\Theta)$. The smoothness condition in (ii) follows immediately from Assumption 2 and the likelihood expressions in the Supplementary Appendix.

Step 2, CLT: The asymptotic normality of the score in (A.6) follows by Theorem 18.10[iv] in van der Vaart (2000) by showing that,

$$\|\ell'_T\big(\boldsymbol{\theta}_0, \boldsymbol{f}_1^{(0:1)}\big) - \ell'_T\big(\boldsymbol{\theta}_0\big)\| \xrightarrow{e.a.s.} 0 \ \text{ as } \ T \to \infty. \tag{A.5}$$

From this, we conclude that $\|\sqrt{T}\ell'_T\big(\boldsymbol{\theta}_0, \boldsymbol{f}_1^{(0:1)}\big) - \sqrt{T}\ell'_T\big(\boldsymbol{\theta}_0\big)\| = \sqrt{T}\|\ell'_T\big(\boldsymbol{\theta}_0, \boldsymbol{f}_1^{(0:1)}\big) - \ell'_T\big(\boldsymbol{\theta}_0\big)\| \xrightarrow{a.s.} 0$ as $T \to \infty$. We apply the CLT for SE martingales in Billingsley (1961) to obtain

$$\sqrt{T}\ell'_T\big(\boldsymbol{\theta}_0\big) \xrightarrow{d} \text{N}\big(0, \mathcal{J}(\boldsymbol{\theta}_0)\big) \ \text{ as } \ T \to \infty, \tag{A.6}$$

where $\mathcal{J}(\boldsymbol{\theta}_0) = \mathbb{E}(\tilde{\ell}'_t\big(\boldsymbol{\theta}_0\big)\tilde{\ell}'_t\big(\boldsymbol{\theta}_0\big)^\top) < \infty$, where finite (co)variances follow from the assumption $n_{\ell'} \geq 2$ in Assumption 7 and the expressions for the likelihood in Section B.1 of the Supplementary Appendix.

To establish the e.a.s. convergence in (A.5), we use the e.a.s. convergence $|f_t(y^{1:t-1}, \boldsymbol{\theta}_0, \bar{f}) - f_t(y^{t-1}, \boldsymbol{\theta}_0)| \xrightarrow{e.a.s.} 0$ and

$$\|\boldsymbol{f}_t^{(1)}(y^{1:t-1}, \boldsymbol{\theta}_0, \boldsymbol{f}_1^{(0:1)}) - \boldsymbol{f}_t^{(1)}(y^{1:t-1}, \boldsymbol{\theta}_0)\| \xrightarrow{e.a.s.} 0,$$

as implied by Proposition 2 under the maintained assumptions. From the differentiability of

$$\tilde{\ell}'_t(\boldsymbol{\theta}, \boldsymbol{f}_1^{(0:1)}) = \ell'\big(\boldsymbol{\theta}, y^{1:t}, \boldsymbol{f}_t^{(0:1)}(y^{1:t-1}, \boldsymbol{\theta}, \boldsymbol{f}_1^{(0:1)})\big)$$

in $\boldsymbol{f}_t^{(0:1)}(y^{1:t-1}, \boldsymbol{\theta}, \boldsymbol{f}_1^{(0:1)})$ and the convexity of $\mathcal{F}$, we use the mean-value theorem to obtain

$$\|\ell_T'(\boldsymbol{\theta}_0,\boldsymbol{f}_1^{(0:1)})-\ell_T'(\boldsymbol{\theta}_0)\| \leq \sum_{j=1}^{4+d_\lambda}\Big|\frac{\partial\ell'(y^{1:t},\hat{\boldsymbol{f}}_t^{(0:1)})}{\partial f_j}\Big| \times \big|\boldsymbol{f}_{j,t}^{(0:1)}(y^{1:t-1},\boldsymbol{\theta}_0,\boldsymbol{f}_1^{(0:1)})-\boldsymbol{f}_{j,t}^{(0:1)}(y^{1:t-1},\boldsymbol{\theta}_0)\big|, \tag{A.7}$$

where $\boldsymbol{f}_{j,t}^{(0:1)}$ denotes the j-th element of $\boldsymbol{f}_t^{(0:1)}$, and $\hat{\boldsymbol{f}}^{(0:1)}$ is on the segment connecting $\boldsymbol{f}_{j,t}^{(0:1)}(y^{1:t-1},\boldsymbol{\theta}_0,\boldsymbol{f}_1^{(0:1)})$ and $\boldsymbol{f}_{j,t}^{(0:1)}$ Note that $\boldsymbol{f}_t^{(0:1)}\in\mathbb{R}^{4+d_\lambda}$ because it contains $f_t\in\mathbb{R}$ as well as $\boldsymbol{f}_t^{(1)}\in\mathbb{R}^{3+d_\lambda}$. Using the expressions of the likelihood and its derivatives, the moment bounds and the moment preserving properties in Assumption 7, Lemma SA.6 in the Supplementary Appendix shows that $\big|\partial\ell'(y^{1:t},\hat{\boldsymbol{f}}_t^{(0:1)})/\partial f\big| = O_p(1)$. The strong convergence in (A.7) is now ensured by

$$\|\ell_T'(\boldsymbol{\theta}_0,\boldsymbol{f}_1^{(0:1)})-\ell_T'(\boldsymbol{\theta}_0)\| = \sum_{i=1}^{4+d_\lambda} O_p(1)o_{e.a.s}(1) = o_{e.a.s.}(1). \tag{A.8}$$

Step 3, uniform convergence of ℓ'': The proof of the uniform convergence in (iv) is similar to that of Theorem 1. We note

$$\sup_{\boldsymbol{\theta}\in\Theta}\|\ell_T''(\boldsymbol{\theta},\bar{f})-\ell_\infty''(\boldsymbol{\theta})\| \leq \sup_{\boldsymbol{\theta}\in\Theta}\|\ell_T''(\boldsymbol{\theta},\bar{f})-\ell_T''(\boldsymbol{\theta})\| + \sup_{\boldsymbol{\theta}\in\Theta}\|\ell_T''(\boldsymbol{\theta})-\ell_\infty''(\boldsymbol{\theta})\|. \tag{A.9}$$

To prove that the first term vanishes a.s., we show that $\sup_{\boldsymbol{\theta}\in\Theta}\|\tilde{\ell}_t''(\boldsymbol{\theta},\bar{f})-\tilde{\ell}_t''(\boldsymbol{\theta})\| \overset{a.s.}{\to} 0$ as $t\to\infty$. The differentiability of $\tilde{g}$, $\tilde{g}'$, $\tilde{p}$, and S from Assumption 2 ensure that $\tilde{\ell}_t''(\cdot,\bar{f})=\ell''(y_t,\boldsymbol{f}_t^{(0:2)}(y^{1:t-1},\cdot,\boldsymbol{f}_{0:2}),\cdot)$ is continuous in $(y_t,\boldsymbol{f}_t^{(0:2)}(y^{1:t-1},\cdot,\boldsymbol{f}_{0:2}))$. Moreover, since all the assumptions of Proposition 2 are satisfied (in particular notice that $s\in\mathbb{C}^{(2,0,2)}(\mathcal{Y}\times\mathcal{F}\times\Lambda)$ is implied by $\tilde{g}\in\mathbb{C}^{(2,0)}(\mathcal{F}\times\mathcal{Y})$, $\tilde{p}\in\mathbb{C}^{(2,2)}(\mathcal{G}\times\Lambda)$ and $S\in\mathbb{C}^{(2,2)}(\mathcal{F}\times\Lambda)$), there exists a unique SE sequence $\{\boldsymbol{f}_t^{(0:2)}(y^{t-1},\cdot)\}_{t\in\mathbb{Z}}$ with elements taking values in $\mathbb{C}(\Theta\times\mathcal{F}^{(0:i)})$ such that $\sup_{\boldsymbol{\theta}\in\Theta}\big\|(y_t,\boldsymbol{f}_t^{(0:2)}(y^{1:t-1},\boldsymbol{\theta},\boldsymbol{f}_{0:2}))-(y_t,\boldsymbol{f}_t^{(0:2)}(y^{t-1},\boldsymbol{\theta})\big\| \overset{a.s.}{\to} 0$ and satisfying, for for $n_f\geq 1$, $\sup_t\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\|\boldsymbol{f}_t^{(0:2)}(y^{1:t-1},\boldsymbol{\theta},\boldsymbol{f}_{0:2})\|^{n_f}<\infty$ and also $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\|\boldsymbol{f}_t^{(0:2)}(y^{t-1},\boldsymbol{\theta})\|^{n_f}<\infty$. The first term in (A.9) now converges to 0 (a.s.) by an application of a continuous mapping theorem for $\ell'':\mathbb{C}(\Theta\times\mathcal{F}^{(0:2)})\to\mathbb{R}$.

The second term in (A.9) converges under a bound $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\|\tilde{\ell}_t''(\boldsymbol{\theta})\|<\infty$ by the SE nature of $\{\ell_T''\}_{t\in\mathbb{Z}}$. The latter is implied by continuity of ℓ'' on the SE sequence $\{(y_t,\boldsymbol{f}_t^{(0:2)}(y^{1:t-1},\cdot))\}_{t\in\mathbb{Z}}$ and Proposition 4.3 in Krengel (1985), where

SE of $\{(y_t, \boldsymbol{f}_t^{(0:2)}(y^{1:t-1}, \cdot))\}_{t\in\mathbb{Z}}$ follows from Proposition 2 under the maintained assumptions. The moment bound $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} \|\tilde{\ell}_t''(\boldsymbol{\theta})\| < \infty$ follows from $n_{\ell''} \geq 1$ in Assumption 7 and Lemma SA.5 in the Supplementary Appendix.

Finally, the non-singularity of the limit $\ell_\infty''(\boldsymbol{\theta}) = \mathbb{E}\tilde{\ell}_t''(\boldsymbol{\theta}) = \mathcal{I}(\boldsymbol{\theta})$ in (v) is implied by the uniqueness of $\boldsymbol{\theta}_0$ as a maximum of $\ell_\infty''(\boldsymbol{\theta})$ in Θ and the usual *second derivative test* calculus theorem. □

Proof of Corollary 2. The desired result is obtained by applying Corollary 1 to guarantee that under the maintained assumptions, $\{y_t\}_{t\in\mathbb{Z}} \equiv \{y_t(\boldsymbol{\theta}_0)\}_{t\in\mathbb{Z}}$ is an SE sequence satisfying $\mathbb{E}|y_t(\boldsymbol{\theta}_{\mathbf{0}})|^{n_y} < \infty$, that $\boldsymbol{\theta}_0 \in \Theta$ be the unique maximizer of $\ell_\infty(\boldsymbol{\theta}, \bar{f})$ on Θ, and then following the same argument as in the proof of Theorem 4. □

Acknowledgements

We thank Peter Boswijk, Christian Francq, Andrew Harvey, and Anders Rahbek, as well as the participants of the "2013 Workshop on Dynamic Models driven by the Score of Predictive Likelihoods", Amsterdam; the "7th International Conference on Computational and Financial Econometrics", London; and the "2014 Workshop on Dynamic Models driven by the Score of Predictive Likelihoods", Tenerife, for helpful comments and discussions.

References

Andres, P. (2014). Computation of maximum likelihood estimates for score driven models for positive valued observations. *Computational Statistics and Data Analysis*, forthcoming.

Andrews, D. W. (1992). Generic uniform convergence. *Econometric Theory 8*, 241–257.

Bauwens, L. and P. Giot (2000). The logarithmic acd model: an application to the bid-ask quote process of three nyse stocks. *Annales d'Economie et de Statistique 60*, 117–149.

Billingsley, P. (1961). The lindeberg-levy theorem for martingales. *Proceedings of the American Mathematical Society 12*(5), 788–792.

Blasques, F., S. J. Koopman, and A. Lucas (2012). Stationarity and ergodicity of univariate generalized autoregressive score processes. *Discussion Paper 12-059, Tinbergen Institute*.

Blasques, F., S. J. Koopman, and A. Lucas (2014a). Maximum likelihood estimation for correctly specified generalized autoregressive score models. *Mimeo, VU University Amsterdam*.

Blasques, F., S. J. Koopman, and A. Lucas (2014b). Supplemental appendix to: Maximum likelihood estimation for generalized autoregressive score models. *VU University Amsterdam*.

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

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

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 (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics 28*(5), 777–795.

Creal, D., B. Schwaab, S. J. Koopman, and A. Lucas (2014). Observation driven mixed-measurement dynamic factor models. *Review of Economics and Statistics*, forthcoming.

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.

Francq, C. and J.-M. Zakoïan (2010). *GARCH Models: Structure, Statistical Inference and Financial Applications*. Wiley.

Gallant, R. and H. White (1988). *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*. Cambridge University Press.

Grammig, J. and K. O. Maurer (2000). Non-monotonic hazard functions and the autoregressive conditional duration model. *Econometrics Journal 3*, 16–38.

Granger, C. and A. Andersen (1978). On the invertibility of time series models. *Stochastic Processes and their Applications 8*(1), 87 – 92.

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

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

Krengel, U. (1985). *Ergodic theorems*. Berlin: De Gruyter studies in Mathematics.

Lee, S. W. and B. E. Hansen (1994). Asymptotic theory for the garch(1,1) quasi-maximum likelihood estimator. *Econometric Theory 10*, 29–52.

Lucas, A., B. Schwaab, and X. Zhang (2014). Conditional euro area sovereign default risk. *Journal of Business and Economic Statistics*, forthcoming.

Lumsdaine, R. L. (1996). Consistency and normality of the quasi-maximum likelihood estimator in igarch(1,1) and covariance stationary garch(1,1) models. *Econometrica 64*, 575–596.

Nelson, D. B. (1990). Arch models as diffusion approximations. *Journal of Econometrics 45*(1-2), 7–38.

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

Nelson, D. B. and D. P. Foster (1994). Asymptotic filtering theory for univariate ARCH models. *Econometrica*, 1–41.

Oh, D. H. and A. Patton (2013). Time-varying systemic risk: Evidence from a dynamic copula model of cds spreads. *Duke University Discussion Paper*.

O'Hagan, A. and T. Leonard (1976). Bayes estimation subject to uncertainty about parameter constraints. *Biometrika 63*(1), 201–203.

Potscher, B. M. and I. R. Prucha (1994). Generic uniform convergence and equicontinuity concepts for random functions: An exploration of the basic structure. *Journal of Econometrics 60*(1-2), 23–63.

Rao, R. R. (1962). Relations between Weak and Uniform Convergence of Measures with Applications. *The Annals of Mathematical Statistics 33*(2), 659–680.

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. (2005). *Estimation in Conditionally Heteroscedastic Time Series Models, Lecture Notes in Statistics 181.* Berlin: Springer-Verlag.

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

van der Vaart, A. W. (2000, June). *Asymptotic Statistics (Cambridge Series in Statistical and Probabilistic Mathematics).* Cambridge University Press.

White, H. (1994). *Estimation, Inference and Specification Analysis.* Cambridge Books. Cambridge University Press.

Wintenberger, O. (2013, January). Continuous invertibility and stable QML estimation of the EGARCH(1,1) model. MPRA Paper 46027, University Library of Munich, Germany.

Supplemental Appendix to: Maximum Likelihood Estimation for Generalized Autoregressive Score Models[1]

Francisco Blasques[a], Siem Jan Koopman[a,b], André Lucas[a]

[(a)] *VU University Amsterdam and Tinbergen Institute*

[(b)] *CREATES, Aarhus University*

B Likelihood and Derivatives Processes

B.1 Explicit expressions for the likelihood and its derivatives

We assume that $\lambda \in \mathbb{R}$. Similar derivations hold for vector valued $\lambda \in \mathbb{R}^{d_\lambda}$. The likelihood function of the GAS model is given by

$$\begin{aligned}
\ell_T(\boldsymbol{\theta}, \bar{f}) &= \frac{1}{T}\sum_{t=2}^{T} \ell_t(\boldsymbol{\theta}, \bar{f}) = \frac{1}{T}\sum_{t=2}^{T} \ell(f_t, y_t; \lambda) \qquad \text{(B.1)} \\
&= \frac{1}{T}\sum_{t=2}^{T} \log p_u\Big(g^{-1}(f_t, y_t)\ ;\ \lambda\Big) + \log \frac{\partial g^{-1}(f_t, y_t)}{\partial y} \\
&= \frac{1}{T}\sum_{t=2}^{T} \log p_u(\tilde{g}_t; \lambda) + \log \frac{\partial \tilde{g}_t}{\partial y} \\
&= \frac{1}{T}\sum_{t=2}^{T} \tilde{p}_t + \log \tilde{g}_t'.
\end{aligned}$$

[1]We thank Peter Boswijk, Christian Francq, Andrew Harvey, and Anders Rahbek, as well as the participants of the "2013 Workshop on Dynamic Models driven by the Score of Predictive Likelihoods", Amsterdam; the "7th International Conference on Computational and Financial Econometrics (CFE 2013)", London; and the "2014 Workshop on Dynamic Models driven by the Score of Predictive Likelihoods", Tenerife, for helpful comments and discussions. Lucas and Blasques thank the Dutch National Science Foundation (NWO; grant VICI453-09-005) for financial support. Email correspondence: f.blasques@vu.nl, s.j.koopman@vu.nl, a.lucas@vu.nl

Its derivative is given by

$$\begin{aligned}
\ell_T'(\boldsymbol{\theta},\bar f) &= \frac{\partial \ell_T(\boldsymbol{\theta},\bar f)}{\partial\boldsymbol{\theta}} = \sum_{t=2}^{T}\ell_t'(\boldsymbol{\theta},\bar f) \qquad \text{(B.2)}\\
&= \sum_{t=2}^{T}\ell_t'\big(y_t,\boldsymbol{f}_t^{(0:1)}(\boldsymbol{\theta},\bar f);\lambda\big) = \frac{1}{T}\sum_{t=2}^{T}\frac{\partial f_t}{\partial\boldsymbol{\theta}}\cdot A_t^* + \frac{\partial\tilde p_t}{\partial\boldsymbol{\theta}},
\end{aligned}$$

with

$$A_t^* := \frac{\partial\tilde p_t}{\partial f_t} + \frac{\partial \log \tilde g_t'}{\partial f_t},$$

and

$$\frac{\partial f_t}{\partial\boldsymbol{\theta}} = \begin{bmatrix}\frac{\partial f_t}{\partial\omega} & \frac{\partial f_t}{\partial\alpha} & \frac{\partial f_t}{\partial\beta} & \frac{\partial f_t}{\partial\lambda}\end{bmatrix}^\top, \qquad \frac{\partial\tilde p_t}{\partial\boldsymbol{\theta}} := \begin{bmatrix}0 & 0 & 0 & \frac{\partial\tilde p_t}{\partial\lambda}\end{bmatrix}^\top.$$

The second derivative is given by

$$\begin{aligned}
\ell_T''(\boldsymbol{\theta},\bar f) &= \frac{\partial^2\ell(\boldsymbol{\theta},\bar f)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} \qquad \text{(B.3)}\\
&= \frac{1}{T}\sum_{t=2}^{T}\frac{\partial^2 f_t}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\cdot A_t^* + \frac{\partial f_t}{\partial\boldsymbol{\theta}}\cdot\frac{\partial A_t^*}{\partial\boldsymbol{\theta}^\top} + \frac{\partial^2\tilde p_t}{\partial\boldsymbol{\theta}\partial f_t}\frac{\partial f_t}{\partial\boldsymbol{\theta}^\top} + \frac{\partial^2\tilde p_t}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\\
&= \frac{1}{T}\sum_{t=2}^{T}\frac{\partial^2 f_t}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\cdot A_t^* + \frac{\partial f_t}{\partial\boldsymbol{\theta}}\frac{\partial f_t}{\partial\boldsymbol{\theta}^\top}\cdot B_t^*\\
&\qquad + \frac{\partial f_t}{\partial\boldsymbol{\theta}}(C_t^*)^\top + C_t^*\frac{\partial f_t}{\partial\boldsymbol{\theta}^\top} + \frac{\partial^2\tilde p_t}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top},
\end{aligned}$$

where

$$\begin{aligned}
B_t^* &= \frac{\partial^2\tilde p_t}{\partial f_t^2} + \frac{\partial^2\log\tilde g_t'}{\partial f_t^2},\\
C_t^* &= \begin{bmatrix}0 & 0 & 0 & \frac{\partial^2\tilde p_t}{\partial f_t\partial\lambda}\end{bmatrix}^\top,\\
\frac{\partial^2 f_t}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} &= \begin{bmatrix}
\frac{\partial^2 f_t}{\partial\omega^2} & \frac{\partial^2 f_t}{\partial\omega\partial\alpha} & \frac{\partial^2 f_t}{\partial\omega\partial\beta} & \frac{\partial^2 f_t}{\partial\omega\partial\lambda}\\
\frac{\partial^2 f_t}{\partial\alpha\partial\omega} & \frac{\partial^2 f_t}{\partial\alpha^2} & \frac{\partial^2 f_t}{\partial\alpha\partial\beta} & \frac{\partial^2 f_t}{\partial\alpha\partial\lambda}\\
\frac{\partial^2 f_t}{\partial\beta\partial\omega} & \frac{\partial^2 f_t}{\partial\beta\partial\alpha} & \frac{\partial^2 f_t}{\partial\beta^2} & \frac{\partial^2 f_t}{\partial\beta\partial\lambda}\\
\frac{\partial^2 f_t}{\partial\lambda\partial\omega} & \frac{\partial^2 f_t}{\partial\lambda\partial\alpha} & \frac{\partial^2 f_t}{\partial\lambda\partial\beta} & \frac{\partial^2 f_t}{\partial\lambda^2}
\end{bmatrix},\\
\frac{\partial^2\tilde p_t}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} &= \begin{bmatrix}
0 & 0 & 0 & 0\\
0 & 0 & 0 & 0\\
0 & 0 & 0 & 0\\
0 & 0 & 0 & \frac{\partial^2\tilde p_t}{\partial\lambda^2}
\end{bmatrix}.
\end{aligned}$$

B.2 Expressions for the derivative processes of f_t

We have $\boldsymbol{\theta} = (\omega,\alpha,\beta,\lambda) \in \Theta$ and write $\partial s(f_t,v_t;\lambda)/\partial\boldsymbol{\theta}_i$ as the derivative of the scaled score w.r.t. λ only. Differentiating the GAS transition equation, we obtain

$$
\begin{aligned}
\frac{\partial f_{t+1}}{\partial \boldsymbol{\theta}_i} &= \frac{\partial \omega}{\partial \boldsymbol{\theta}_i} + \frac{\partial \alpha}{\partial \boldsymbol{\theta}_i} s_t + \alpha \frac{\partial s_t}{\partial f_t} \frac{\partial f_t}{\partial \boldsymbol{\theta}_i} + \alpha \frac{\partial s_t}{\partial \boldsymbol{\theta}_i} + \frac{\partial \beta}{\partial \boldsymbol{\theta}_i} f_t + \beta \frac{\partial f_t}{\partial \boldsymbol{\theta}_i}, \\
&= \boldsymbol{A}_t^{(1)} + \boldsymbol{B}_t \frac{\partial f_t}{\partial \boldsymbol{\theta}_i},
\end{aligned}
$$

with

$$
\begin{aligned}
\boldsymbol{A}_t^{(1)} &= \boldsymbol{A}_t^{(1)}(\boldsymbol{\theta}) = \frac{\partial \omega}{\partial \boldsymbol{\theta}} + \frac{\partial \alpha}{\partial \boldsymbol{\theta}} s_t + \alpha \frac{\partial s_t}{\partial \boldsymbol{\theta}} + \frac{\partial \beta}{\partial \boldsymbol{\theta}} f_t, \\
\boldsymbol{B}_t &= \boldsymbol{B}_t(\boldsymbol{\theta}) = \alpha \frac{\partial s_t}{\partial f_t} + \beta.
\end{aligned}
$$

Similarly, we obtain a recursion for the second derivative process

$$
\begin{aligned}
\frac{\partial^2 f_{t+1}}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} &= \frac{\partial \boldsymbol{A}_t^{(1)}}{\partial \boldsymbol{\theta}^\top} + \frac{\partial \boldsymbol{A}_t^{(1)}}{\partial f_t} \frac{\partial f_t}{\partial \boldsymbol{\theta}^\top} + \frac{\partial f_t}{\partial \boldsymbol{\theta}} \frac{\partial \boldsymbol{B}_t}{\partial \boldsymbol{\theta}^\top} + \frac{\partial \boldsymbol{B}_t}{\partial f_t} \frac{\partial f_t}{\partial \boldsymbol{\theta}} \frac{\partial f_t}{\partial \boldsymbol{\theta}^\top} + \boldsymbol{B}_t \frac{\partial^2 f_t}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \\
&= \boldsymbol{A}_t^{(2)} + \boldsymbol{B}_t \frac{\partial^2 f_t}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top},
\end{aligned}
$$

with

$$
\begin{aligned}
\boldsymbol{A}_t^{(2)} &= \frac{\partial \boldsymbol{A}_t^{(1)}}{\partial \boldsymbol{\theta}^\top} + \frac{\partial \boldsymbol{A}_t^{(1)}}{\partial f_t} \frac{\partial f_t}{\partial \boldsymbol{\theta}^\top} + \frac{\partial f_t}{\partial \boldsymbol{\theta}} \frac{\partial \boldsymbol{B}_t}{\partial \boldsymbol{\theta}^\top} + \frac{\partial \boldsymbol{B}_t}{\partial f_t} \frac{\partial f_t}{\partial \boldsymbol{\theta}} \frac{\partial f_t}{\partial \boldsymbol{\theta}^\top} \\
&= \frac{\partial \alpha}{\partial \boldsymbol{\theta}} \frac{\partial s_t}{\partial \boldsymbol{\theta}^\top} + \frac{\partial s_t}{\partial \boldsymbol{\theta}} \frac{\partial \alpha}{\partial \boldsymbol{\theta}^\top} + \alpha \frac{\partial^2 s_t}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} + \Big(\frac{\partial \alpha}{\partial \boldsymbol{\theta}} \frac{\partial s_t}{\partial f_t} + \alpha \frac{\partial^2 s_t}{\partial \boldsymbol{\theta} \partial f_t} + \frac{\partial \beta}{\partial \boldsymbol{\theta}} \Big) \frac{\partial f_t}{\partial \boldsymbol{\theta}^\top} \\
&\quad + \frac{\partial f_t}{\partial \boldsymbol{\theta}^\top} \Big(\frac{\partial s_t}{\partial f_t} \frac{\partial \alpha}{\partial \boldsymbol{\theta}^\top} + \alpha \frac{\partial^2 s_t}{\partial f_t \partial \boldsymbol{\theta}^\top} + \frac{\partial \beta}{\partial \boldsymbol{\theta}^\top} \Big) + \alpha \frac{\partial^2 s_t}{\partial f_t^2} \frac{\partial f_t}{\partial \boldsymbol{\theta}} \frac{\partial f_t}{\partial \boldsymbol{\theta}^\top}.
\end{aligned}
$$

C Properties and Moments of Stochastic Processes from Stochastic Recurrence Equations

Propositions SA.1 and SA.2 below are written for general random sequences $\{x_t(v_{\boldsymbol{\theta}}^{1:t-1}, \boldsymbol{\theta}, \bar{x})\}_{t \in \mathbb{N}}$ taking values in $\mathcal{X} \subseteq \mathbb{R}$, where $x_t(v_{\boldsymbol{\theta}}^{1:t-1}, \boldsymbol{\theta}, \bar{x})$ is generated by a stochastic recurrence equation of the form

$$
x_{t+1}(v_{\boldsymbol{\theta}}^{1:t}, \boldsymbol{\theta}, \bar{x}) = \phi\big(x_t(v_{\boldsymbol{\theta}}^{1:t-1}, \boldsymbol{\theta}, \bar{x}), v_t(\boldsymbol{\theta}), \boldsymbol{\theta}\big), \tag{C.1}
$$

where $\bar{x} \in \mathcal{X}$ is a fixed initialization value at $t = 1$, $\phi : \mathcal{X} \times \mathcal{V} \times \Theta \to \mathcal{X}$ is a continuous map, $\mathcal{X}$ is a convex set $\mathcal{X} \subseteq \mathcal{X}^* \subseteq \mathbb{R}$, and $\boldsymbol{\theta} \in \Theta$ is a static parameter vector. For the results that follow we define the supremum of the kth power random Lipschitz constant as

$$
r_t^k(\boldsymbol{\theta}) := \sup_{(x,x') \in \mathcal{X}^* \times \mathcal{X}^* : x \neq x'} \frac{|\phi(x, v_t(\boldsymbol{\theta}), \boldsymbol{\theta}) - \phi(x', v_t(\boldsymbol{\theta}), \boldsymbol{\theta})|^k}{|x - x'|^k}, \quad k \geq 1.
$$

Moreover, for random sequences $\{x_{1,t}\}_{t\in\mathbb{Z}}$ and $\{x_{2,t}\}_{t\in\mathbb{Z}}$, we say that $x_{1,t}$ converges exponentially fast almost surely (e.a.s.) to $x_{2,t}$ if there exists a constant $c > 1$ such that $c^t \|x_{1,t} - x_{2,t}\| \overset{a.s.}{\to} 0$; see also SM06.

Proposition SA.1. *For every $\boldsymbol{\theta} \in \Theta$, let $\{v_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ be a strictly stationary and ergodic (SE) sequence and assume $\exists\ \bar{x} \in \mathcal{X}$ such that*

(i) $\mathbb{E}\log^+ |\phi(\bar{x}, v_1(\boldsymbol{\theta}), \boldsymbol{\theta}) - \bar{x}| < \infty$;

(ii) $\mathbb{E}\ \log r_1^1(\boldsymbol{\theta}) < 0$.

Then $\{x_t(v_{\boldsymbol{\theta}}^{1:t-1}, \boldsymbol{\theta}, \bar{x})\}_{t\in\mathbb{N}}$ converges e.a.s. to a unique SE solution $\{x_t(v_{\boldsymbol{\theta}}^{t-1}, \boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ for every $\boldsymbol{\theta} \in \Theta$ as $t \to \infty$.

If furthermore, for every $\boldsymbol{\theta} \in \Theta\ \exists\ n > 0$ such that

(iii.a) $\|\phi(\bar{x}, v_1(\boldsymbol{\theta}), \boldsymbol{\theta})\|_n < \infty$;

(iv.a) $\mathbb{E}r_1^n(\boldsymbol{\theta}) < 1$;

(v.a) $x_t(v_{\boldsymbol{\theta}}^{1:t-1}, \boldsymbol{\theta}, \bar{x}) \perp r_t^n(\boldsymbol{\theta})\ \forall\ (t, \bar{x}) \in\ \mathbb{N} \times \mathcal{X}$;

then $\sup_t \mathbb{E}|x_t(v_{\boldsymbol{\theta}}^{1:t-1}, \boldsymbol{\theta}, \bar{x})|^n < \infty$ and $\mathbb{E}|x_t(v_{\boldsymbol{\theta}}^{t-1}, \boldsymbol{\theta})|^n < \infty\ \forall\ \boldsymbol{\theta} \in \Theta$.

Alternatively, if instead of (iii.a)-(v.a) we have for every $\boldsymbol{\theta} \in \Theta$

(iii.b) $|\bar{\phi}(\bar{x}, \boldsymbol{\theta})| := \sup_{v\in\mathcal{V}} |\phi(\bar{x}, v, \boldsymbol{\theta})| < \infty$;

(iv.b) $\sup_{(x,x')\in\mathcal{X}\times\mathcal{X}:x\neq x'} |\bar{\phi}(x, \boldsymbol{\theta}) - \bar{\phi}(x', \boldsymbol{\theta})|/|x - x'| < 1$;

then $\sup_t \mathbb{E}|x_t(v_{\boldsymbol{\theta}}^{1:t-1}, \boldsymbol{\theta}, \bar{x})|^n < \infty$ and $\mathbb{E}|x_t(v_{\boldsymbol{\theta}}^{t-1}, \boldsymbol{\theta})|^n < \infty$ holds for all $\boldsymbol{\theta} \in \Theta$ and every $n > 0$.

Proposition SA.1 not only establishes the convergence to a unique SE solution, but also establishes the existence of unconditional moments. The latter property is key to proving the consistency and asymptotic normality of the MLE in Section 4 of the paper. To establish convergence to an SE solution, condition *(ii)* requires the stochastic recurrence equation to be contracting on average. For the subsequent existence of moments, the contraction condition *(iv.a)*, together with the moment bound in *(iii.a)*, and the independence assumption *(v.a)*, are sufficient. Alternatively, if by condition *(iii.b)* ϕ is uniformly bounded in v, then a deterministic contraction condition *(iv.b)* only needs to hold on the uniform bound and the moment bound holds for any $n \geq 1$. Note that conditions *(i)–(ii)* are implied by *(iii.a)–(v.a)*. Remark SA.1 shows that condition *(v.a)* is automatically satisfied if $v_t(\boldsymbol{\theta})$ is an innovation sequence.

Remark SA.1. If $v_t(\boldsymbol{\theta}) \perp x_t(v_{\boldsymbol{\theta}}^{1:t-1}, \boldsymbol{\theta}, \bar{x})$ then *(v.a)* in Proposition SA.1 holds.

The condition that $v_t(\boldsymbol{\theta})$ is an innovation sequence is typically more intuitive. We keep the independence assumption *(v.a)* in Proposition SA.1, however, because in some of our models the supremum Lipschitz constant is independent of the random $v_t(\boldsymbol{\theta})$. In such cases, the independence is easily satisfied, even in cases where $v_t(\boldsymbol{\theta})$ is not an innovation sequence.

Following SM06, we also note that conditions *(i)* and *(ii)* in Proposition SA.1 provide us with an almost sure representation of $x_t(v_{\boldsymbol{\theta}}^{1:t-1}, \boldsymbol{\theta}, \bar{x})$ as a measurable function of $v_{\boldsymbol{\theta}}^{1:t-1}$.

Remark SA.2. Let conditions *(i)* and *(ii)* of Proposition SA.1 hold. Then $x_t(v_{\boldsymbol{\theta}}^{t-1}, \boldsymbol{\theta})$ admits the following a.s. representation for every $\boldsymbol{\theta} \in \Theta$

$$x_t(v_{\boldsymbol{\theta}}^{t-1}, \boldsymbol{\theta}) = \lim_{r\to\infty} \phi\big(\cdot, v_{t-1}(\boldsymbol{\theta}), \boldsymbol{\theta}\big) \circ \phi\big(\cdot, v_{t-2}(\boldsymbol{\theta}), \boldsymbol{\theta}\big) \circ ... \circ \phi\big(\cdot, v_{t-r}(\boldsymbol{\theta}), \boldsymbol{\theta}\big),$$

and $x_t(v_{\boldsymbol{\theta}}^{t-1}, \boldsymbol{\theta})$ is measurable with respect to the σ-algebra generated by $v_{\boldsymbol{\theta}}^{t-1}$.

Remark SA.3. *Conditions (iii)–(v) in Proposition 1 of the paper can be substituted by*

$$\sup_{u\in\mathcal{U}} |s_u(\bar{f}^*, u; \lambda)| = |\bar{s}_u(\bar{f}^*; \lambda)| < \infty \quad \textit{and} \quad \sup_{f^*\in\mathcal{F}} |\partial s_u(f^*; \lambda)/\partial f| < 1.$$

Proposition SA.2 deals with sequences $\{x_t(v^{1:t-1}, \boldsymbol{\theta}, \bar{x})\}_{t\in\mathbb{N}}$ that, for a given initialization $\bar{x} \in \mathcal{X}$, are generated by

$$x_{t+1}(v^{1:t}, \boldsymbol{\theta}, \bar{x}) = \phi\big(x_t(v^{1:t-1}, \boldsymbol{\theta}, \bar{x}), v_t, \boldsymbol{\theta}\big) \ \forall \ (\boldsymbol{\theta}, t) \in \Theta \times \mathbb{N},$$

where $\phi : \mathcal{X} \times \mathcal{V} \times \Theta \to \mathcal{X}$ is continuous, and where $v^{1:t}$ now replaces $v_{\boldsymbol{\theta}}^{1:t}$. We have the following proposition.

Proposition SA.2. *Let Θ be compact, $\{v_t\}_{t\in\mathbb{Z}}$ be stationary and ergodic (SE) and assume there exists an $\bar{x} \in \mathcal{X}$, such that*

(i) $\mathbb{E}\log^+ \sup_{\boldsymbol{\theta}\in\Theta} |\phi(\bar{x}, v_t, \boldsymbol{\theta}) - \bar{x}| < \infty$;

(ii) $\mathbb{E}\log \sup_{\boldsymbol{\theta}\in\Theta} r_1^1(\boldsymbol{\theta}) < 0$.

Then $\{x_t(v^{1:t-1}, \boldsymbol{\theta}, \bar{x})\}_{t\in\mathbb{N}}$ converges e.a.s. to a unique SE solution $\{x_t(v^{t-1}, \boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ uniformly on Θ as $t \to \infty$.

If furthermore $\exists\ n > 0$ such that either

(iii.a) $\|\phi(\bar{x}, v_t, \cdot)\|_n^{\Theta} < \infty$;

(iv.a) $\sup_{\boldsymbol{\theta}\in\Theta} |\phi(x, v, \boldsymbol{\theta}) - \phi(x', v, \boldsymbol{\theta})| \; < \; |x - x'| \;\; \forall \; (x, x', v) \in \mathcal{X} \times \mathcal{X} \times \mathcal{V}$*;*

or

(iii.b) $\|\phi(\bar{x}, v_t, \cdot)\|_n^{\Theta} < \infty$*;*

(iv.b) $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} r_1^n(\boldsymbol{\theta}) < 1$*;*

(v.b) $x_t(v^{1:t-1}, \boldsymbol{\theta}, \bar{x}) \perp r_t^n(\boldsymbol{\theta}) \; \forall \; (t, \bar{x}, \boldsymbol{\theta}) \in \mathbb{N} \times \mathcal{X} \times \Theta$*;*

then $\sup_t \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |x_t(v^{1:t-1}, \boldsymbol{\theta}, \bar{x})|^n < \infty$ *and* $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |x_t(v^{t-1}, \boldsymbol{\theta})|^n < \infty$*.*

If instead of (iii.a)-(iv.a) or (iii.b)-(v.b) we have

(iii.c) $\sup_{\boldsymbol{\theta}\in\Theta} \sup_{v\in\mathcal{V}} |\phi(\bar{x}, v, \boldsymbol{\theta})| = |\bar{\phi}(\bar{x}, \boldsymbol{\theta})| < \infty$*;*

(iv.c) $\sup_{\boldsymbol{\theta}\in\Theta} \sup_{(x,x')\in\mathcal{X}^*\times\mathcal{X}^*:x\neq x'} |\bar{\phi}(x, \boldsymbol{\theta}) - \bar{\phi}(x', \boldsymbol{\theta})| < |x - x'|$*;*

then $\sup_t \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |x_t(v^{1:t-1}, \boldsymbol{\theta}, \bar{x})|^n < \infty$ *and* $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |x_t(v^{t-1}, \boldsymbol{\theta})|^n < \infty$ *for every* $n > 0$*.*

The contraction condition *(iv.a)* in Proposition SA.2 is stricter than condition *(iv.b)*. Rather than only requiring the contraction property to hold in expectation, condition *(iv.a)* holds for all $v \in \mathcal{V}$.

Again, we note that conditions *(i)* and *(ii)* in Proposition SA.2 provide us with an almost sure representation of $x_t(v^{t-1}, \boldsymbol{\theta})$ in terms of v^{t-1}.

Remark SA.4. Let conditions *(i)* and *(ii)* of Proposition SA.2 hold. Then $x_t(v^{t-1}, \boldsymbol{\theta})$ admits the following a.s. representation for every $\boldsymbol{\theta} \in \Theta$

$$x_t(v^{t-1}, \boldsymbol{\theta}) = \lim_{r\to\infty} \phi\big(\cdot, v_{t-1}, \boldsymbol{\theta}\big) \circ \phi\big(\cdot, v_{t-2}, \boldsymbol{\theta}\big) \circ ... \circ \phi\big(\cdot, v_{t-r}, \boldsymbol{\theta}\big)$$

and $x_t(v^{t-1}, \boldsymbol{\theta})$ is measurable with respect to the σ-algebra generated by v^{t-1}.

D Proofs of Propositions

Proof of Proposition SA.1. Step 1, SE: The assumption that $\{v_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ is SE $\forall \; \boldsymbol{\theta} \in \Theta$ together with the continuity of ϕ on $\mathcal{X} \times \mathcal{V} \times \Theta$ (and resulting measurability w.r.t. the Borel σ-algebra) implies that $\{\phi_t := \phi(\cdot, v_t(\boldsymbol{\theta}), \boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ is SE for every $\boldsymbol{\theta} \in \Theta$ by Krengel (1985, Proposition 4.3). Condition C1 in Bougerol (1993, Theorem 3.1) is immediately implied by assumption *(i.a)* for every $\boldsymbol{\theta} \in \Theta$.

Condition C2 in Bougerol (1993, Theorem 3.1) is implied, for every $\boldsymbol{\theta} \in \Theta$, by condition *(ii.a)* since for every $\boldsymbol{\theta} \in \Theta$,

$$\mathbb{E} \log \sup_{(x,x') \in \mathcal{X} \times \mathcal{X} : x \neq x'} \frac{|\phi(v_t(\boldsymbol{\theta}), x, \boldsymbol{\theta}) - \phi(v_t(\boldsymbol{\theta}), x', \boldsymbol{\theta})|}{|x - x'|} = \mathbb{E} \log r_t^1(\boldsymbol{\theta}) < 0.$$

As a result, for every $\boldsymbol{\theta} \in \Theta$, $\{x_t(v_{\boldsymbol{\theta}}^{1:t-1}, \boldsymbol{\theta}, \bar{x})\}_{t \in \mathbb{N}}$ converges to an SE solution $\{x_t(v_{\boldsymbol{\theta}}^{t-1}, \boldsymbol{\theta})\}_{t \in \mathbb{Z}}$. Uniqueness and e.a.s. convergence is obtained by Straumann and Mikosch (2006, Theorem 2.8).

Step 2, moment bounds: For $n \geq 1$ the moment bounds are obtained by first noting that for every $\boldsymbol{\theta} \in \Theta$ we have $\sup_t \mathbb{E}|x_t(v_{\boldsymbol{\theta}}^{1:t-1}, \boldsymbol{\theta}, \bar{x})|^n < \infty$ if and only if $\|x_t(v_{\boldsymbol{\theta}}^{1:t-1}, \boldsymbol{\theta}, \bar{x})\|_n < \infty$. Let $\bar{x}_{\boldsymbol{\theta}} = \phi(\bar{x}, \bar{v}, \boldsymbol{\theta})$ for some $\bar{v} \in \mathcal{V}$, then we have for every $\boldsymbol{\theta} \in \Theta$

$$\begin{aligned}
&\sup_t \|x_t(v_{\boldsymbol{\theta}}^{1:t-1}, \boldsymbol{\theta}, \bar{x}) - \bar{x}_{\boldsymbol{\theta}}\|_n && \text{(D.1)}\\
&= \sup_t \|\phi(x_{t-1}(v_{\boldsymbol{\theta}}^{1:t-2}, \boldsymbol{\theta}, \bar{x}), v_{t-1}(\boldsymbol{\theta}), \boldsymbol{\theta}) - \phi(\bar{x}, \bar{v}, \boldsymbol{\theta})\|_n\\
&\leq \sup_t \|\phi(x_{t-1}(v_{\boldsymbol{\theta}}^{1:t-2}, \boldsymbol{\theta}, \bar{x}), v_{t-1}(\boldsymbol{\theta}), \boldsymbol{\theta}) - \phi(\bar{x}, v_{t-1}(\boldsymbol{\theta}), \boldsymbol{\theta})\|_n\\
&\quad + \sup_t \|\phi(\bar{x}, v_{t-1}(\boldsymbol{\theta}), \boldsymbol{\theta})\|_n + |\phi(\bar{x}, \bar{v}, \boldsymbol{\theta})|\\
&\leq \sup_t \Big(\mathbb{E}|x_{t-1}(v_{\boldsymbol{\theta}}^{1:t-2}, \boldsymbol{\theta}, \bar{x}) - \bar{x}|^n\\
&\quad \times \frac{|\phi(x_{t-1}(v_{\boldsymbol{\theta}}^{1:t-2}, \boldsymbol{\theta}, \bar{x}), v_{t-1}(\boldsymbol{\theta}), \boldsymbol{\theta}) - \phi(\bar{x}, v_{t-1}(\boldsymbol{\theta}), \boldsymbol{\theta})|^n}{|x_{t-1}(v_{\boldsymbol{\theta}}^{1:t-2}, \boldsymbol{\theta}, \bar{x}) - \bar{x}|^n}\Big)^{1/n}\\
&\quad + \sup_t \|\phi(\bar{x}, v_{t-1}(\boldsymbol{\theta}), \boldsymbol{\theta})\|_n + |\bar{x}_{\boldsymbol{\theta}}|\\
&\leq \sup_t \|x_{t-1}(v_{\boldsymbol{\theta}}^{1:t-2}, \boldsymbol{\theta}, \bar{x}) - \bar{x}_{\boldsymbol{\theta}}\|_n \cdot \big(\mathbb{E} r_t^n(\boldsymbol{\theta})\big)^{1/n} && \text{(D.2)}\\
&\quad + \sup_t \|\phi(\bar{x}, v_{t-1}(\boldsymbol{\theta}), \boldsymbol{\theta})\|_n + |\bar{x}_{\boldsymbol{\theta}} - \bar{x}|\big(\mathbb{E} r_t^n(\boldsymbol{\theta})\big)^{1/n} + |\bar{x}_{\boldsymbol{\theta}}|,
\end{aligned}$$

where $(\mathbb{E} r_t^n(\boldsymbol{\theta}))^{1/n} < 1$ by assumption *(iv.a)*. Using this inequality, we can unfold the recursion (D.1)–(D.2) as

$$\begin{aligned}
&\sup_t \|x_t(v_{\boldsymbol{\theta}}^{1:t-1}, \boldsymbol{\theta}, \bar{x}) - \bar{x}_{\boldsymbol{\theta}}\|_n \leq |\bar{x} - \bar{x}_{\boldsymbol{\theta}}|\big(\mathbb{E} r_t^n(\boldsymbol{\theta})\big)^{1/n} +\\
&\sum_{j=0}^{t-2} \big(\mathbb{E} r_t^n(\boldsymbol{\theta})\big)^{j/n} \Big(\sup_t \|\phi(\bar{x}, v_{t-1}(\boldsymbol{\theta}), \boldsymbol{\theta})\|_n + |\bar{x}_{\boldsymbol{\theta}} - \bar{x}|\big(\mathbb{E} r_t^n(\boldsymbol{\theta})\big)^{1/n}\\
&\qquad + |\bar{x}_{\boldsymbol{\theta}}|\Big) \quad \leq \quad |\bar{x} - \bar{x}_{\boldsymbol{\theta}}| +\\
&\sum_{j=1}^{t-2} \big(\mathbb{E} r_t^n(\boldsymbol{\theta})\big)^{j/n} \Big(\sup_t \|\phi(\bar{x}, v_{t-j}(\boldsymbol{\theta}), \boldsymbol{\theta})\|_n + |\bar{x}_{\boldsymbol{\theta}} - \bar{x}| + |\bar{x}_{\boldsymbol{\theta}}|\Big) \quad \leq\\
&\frac{|\bar{x}_{\boldsymbol{\theta}} - \bar{x}| + |\bar{x}_{\boldsymbol{\theta}}| + \sup_t \|\phi(\bar{x}, v_t(\boldsymbol{\theta}), \boldsymbol{\theta})\|_n}{1 - (\mathbb{E} r_t^n(\boldsymbol{\theta}))^{1/n}} + |\bar{x} - \bar{x}_{\boldsymbol{\theta}}| < \infty \quad \forall\, \boldsymbol{\theta} \in \Theta.
\end{aligned}$$

The same result can be obtained using conditions *(iii.b)* and *(iv.b)* by noting that

$$\begin{aligned}\sup_t \|x_t(v_{\boldsymbol{\theta}}^{1:t-1},\boldsymbol{\theta},\bar{x})\|_n &\leq \sup_t \|\bar{\phi}(x_{t-1}(v_{\boldsymbol{\theta}}^{1:t-2},\boldsymbol{\theta},\bar{x}),\boldsymbol{\theta})\|_n \\ &\leq \sup_t \|\bar{\phi}(x_{t-1}(v_{\boldsymbol{\theta}}^{1:t-2},\boldsymbol{\theta},\bar{x}),\boldsymbol{\theta}) - \bar{\phi}(\bar{x},\boldsymbol{\theta})\|_n \;+\; \|\bar{\phi}(\bar{x},\boldsymbol{\theta})\|_n \\ &< \sup_t \|x_{t-1}(v_{\boldsymbol{\theta}}^{1:t-2},\boldsymbol{\theta},\bar{x})\|_n + |\bar{x}| + |\bar{\phi}(\bar{x},\boldsymbol{\theta})|.\end{aligned}$$

As a result, unfolding the recursion renders $\sup_t \|x_t(v_{\boldsymbol{\theta}}^{1:t-1},\boldsymbol{\theta},\bar{x})\|_n < \infty$ by the same argument as above.

For $0 < n < 1$ the function $\|\cdot\|_n$ is only a pseudo-norm as it is not sub-additive. However, the proof still follows by working instead with the metric $\|\cdot\|_n^* := (\|\cdot\|_n)^n$ which is sub-additive. □

Proof of Proposition 1. First set

$$\phi(x_t(v_{\boldsymbol{\theta}}^{1:t-1},\boldsymbol{\theta},\bar{x}),v_t(\boldsymbol{\theta}),\boldsymbol{\theta}) = \omega + \alpha s_u(f_t(u_\lambda^{1:t-1},\boldsymbol{\theta},\bar{f}),u_t;\lambda) + \beta\, f_t(u_\lambda^{1:t-1},\boldsymbol{\theta},\bar{f}),$$

$v_t(\boldsymbol{\theta}) = u_t$, and $x_t(v_{\boldsymbol{\theta}}^{1:t-1},\boldsymbol{\theta},\bar{x}) = f_t(u_\lambda^{1:t-1},\boldsymbol{\theta},\bar{f})$. Note that the assumption that $s_u \in \mathbb{C}^{(1,0,0)}(\mathcal{F}\times\mathcal{U}\times\Lambda)$ as stated above Proposition 1, together with the convexity of $\mathcal{F}$ imply $\phi \in \mathbb{C}^{(1,0,0)}(\mathcal{X}\times\mathcal{V}\times\Theta)$ and that $\mathcal{X}$ is convex. By the mean value theorem, we have

$$\begin{aligned}&\mathbb{E}\sup_{(x,x')\in\mathcal{X}\times\mathcal{X}:x\neq x'} \frac{|\phi(x,v_t(\boldsymbol{\theta}),\boldsymbol{\theta}) - \phi(x',v_t(\boldsymbol{\theta}),\boldsymbol{\theta})|^k}{|x-x'|^k} \quad \leq \\ &\mathbb{E}\sup_{x^*\in\mathcal{X}}\Big|\frac{\partial\phi(x^*,v_t(\boldsymbol{\theta}),\boldsymbol{\theta})}{\partial x}\Big|^k = \mathbb{E}\sup_{f^*\in\mathcal{F}}\Big|\beta+\alpha+\frac{\partial s_u(f^*,v_t(\boldsymbol{\theta}),\boldsymbol{\theta})}{\partial f}\Big|^k \quad \forall\, k\geq 1,\end{aligned}$$

such that conditions *(i)–(v)* now directly imply conditions *(i)–(v.a)* in Proposition SA.1. □

Proof of Proposition SA.2. Step 0, additional notation: Following Straumann and Mikosch (2006, Proposition 3.12), the uniform convergence of the process $\sup_{\boldsymbol{\theta}\in\Theta}|x_t(v^{1:t-1},\boldsymbol{\theta},\bar{x}) - x_t(v^{t-1},\boldsymbol{\theta})| \stackrel{e.a.s.}{\to} 0$ is obtained by appealing to Bougerol (1993, Theorem 3.1) using sequences of random functions $\{x_t(v^{1:t-1},\cdot,\bar{x})\}_{t\in\mathbb{N}}$ rather than sequences of real numbers. This change is subtle in the notation, but important. We refer to SM06 for details.

The elements $x_t(v^{t-1}, \cdot, \bar{x})$ are random functions that take values in the separable Banach[2] space $\mathcal{X}_\Theta \subseteq (\mathbb{C}(\Theta, \mathcal{X}), \|\cdot\|^\Theta)$, where $\|x_t(v^{t-1}, \cdot)\|^\Theta \equiv \sup_{\boldsymbol{\theta}\in\Theta} |x_t(v^{t-1}, \boldsymbol{\theta})|$. The functions $x_t(v^{1:t-1}, \cdot, \bar{x})$ are generated by

$$x_t(v^{1:t}, \cdot, \bar{x}) = \phi^*(x_{t-1}(v^{1:t-1}, \cdot, \bar{x}), v_t, \cdot) \ \forall\ t \in \{2, 3, \dots\},$$

with starting function $x_1(\emptyset, \boldsymbol{\theta}, \bar{x}) = \bar{x} \ \forall\ \boldsymbol{\theta} \in \Theta$, and where $\{\phi^*(\cdot, v_t, \cdot)\}_{t\in\mathbb{Z}}$ is a sequence of stochastic recurrence equations $\phi^*(\cdot, v_t, \cdot) : \mathbb{C}(\Theta) \times \Theta \to \mathbb{C}(\Theta) \ \forall\ t$ as in Straumann and Mikosch (2006, Proposition 3.12).

Step 1, SE: With the above notation in place, we now first prove the SE part of the proposition. The assumption that $\{v_t\}_{t\in\mathbb{Z}}$ is SE together with the continuity of ϕ on $\mathcal{X} \times \mathcal{V} \times \Theta$ implies that $\{\phi^*(\cdot, v_t, \cdot)\}_{t\in\mathbb{Z}}$ is SE. Condition C1 in Bougerol (1993, Theorem 3.1) is now implied directly by condition *(i)*, since there exists a function $\bar{x}_\Theta \in \mathbb{C}(\Theta)$ with $\bar{x}_\Theta(\boldsymbol{\theta}) = \bar{x} \ \forall\ \boldsymbol{\theta} \in \Theta$ that satisfies $\mathbb{E}\log^+ \|\phi^*(\bar{x}_\Theta(\cdot), v_t, \cdot) - \bar{x}_\Theta(\cdot)\|^\Theta = \mathbb{E}\log^+ \sup_{\boldsymbol{\theta}\in\Theta} |\phi(\bar{x}, v_t, \boldsymbol{\theta}) - \bar{x}| < \infty$.

Condition C2 in Bougerol (1993, Theorem 3.1) is directly implied by condition *(ii)*, since

$$\begin{aligned}
&\mathbb{E}\log \sup_{(\bar{x}_\Theta, \bar{x}'_\Theta)\in\mathcal{X}_\Theta\times\mathcal{X}_\Theta : \|\bar{x}_\Theta - \bar{x}'_\Theta\|^\Theta > 0} \frac{\|\phi^*(\bar{x}_\Theta(\cdot), v_t, \cdot) - \phi^*(\bar{x}'_\Theta(\cdot), v_t, \cdot)\|^\Theta}{\|\bar{x}_\Theta(\cdot) - \bar{x}'_\Theta(\cdot)\|^\Theta} \\
&= \mathbb{E}\log \sup_{(\bar{x}, \bar{x}')\in\mathcal{X}\times\mathcal{X} : \bar{x}\neq\bar{x}'} \frac{\sup_{\boldsymbol{\theta}\in\Theta} |\phi(\bar{x}, v_t, \boldsymbol{\theta}) - \phi(\bar{x}', v_t, \boldsymbol{\theta})|}{|\bar{x} - \bar{x}'|} \\
&= \mathbb{E}\log \sup_{(\bar{x}, \bar{x}')\in\mathcal{X}\times\mathcal{X} : \bar{x}\neq\bar{x}'} \frac{\sup_{\boldsymbol{\theta}\in\Theta} |\phi(\bar{x}, v_t, \boldsymbol{\theta}) - \phi(\bar{x}', v_t, \boldsymbol{\theta})| \ \sup_{\boldsymbol{\theta}\in\Theta} |\bar{x} - \bar{x}|}{|\bar{x} - \bar{x}'|} \\
&\leq \mathbb{E}\log \sup_{\boldsymbol{\theta}\in\Theta} r_t^1(\boldsymbol{\theta}) < 0.
\end{aligned}$$

As a result, $\{x_t(v^{1:t-1}, \cdot, \bar{x})\}_{t\in\mathbb{N}}$ converges to an SE solution $\{x_t(v^{t-1}, \cdot)\}_{t\in\mathbb{Z}}$ in $\|\cdot\|^\Theta$-norm. Uniqueness and e.a.s. convergence is obtained in Straumann and Mikosch (2006, Theorem 2.8), such that $\sup_{\boldsymbol{\theta}\in\Theta} |x_t(v^{1:t-1}, \boldsymbol{\theta}, \bar{x}) - x_t(v^{t-1}, \boldsymbol{\theta})| \overset{e.a.s.}{\to} 0$.

Step 2, moment bounds: For $n \geq 1$ we use a similar argument as in the proof of Proposition SA.1. We first note that $\sup_t \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |x_t(v^{1:t-1}, \boldsymbol{\theta}, \bar{x})|^n < \infty$ if and only if $\sup_t \|x_t(v^{1:t-1}, \boldsymbol{\theta}, \bar{x})\|_n^\Theta < \infty$. Further, $\|x_t(v^{1:t-1}, \cdot, \bar{x}) - \bar{x}_\Theta\|_n^\Theta < \infty$ implies $\|x_t(v^{1:t-1}, \cdot, \bar{x})\|_n^\Theta < \infty$ for any $\bar{x}_\Theta \in \mathcal{X}_\Theta \subseteq \mathbb{C}(\Theta)$, since continuity on the compact Θ implies $\sup_{\boldsymbol{\theta}\in\Theta} |\bar{x}_\Theta(\boldsymbol{\theta})| < \infty$. For a pair $(\bar{x}, \bar{v}) \in \mathcal{X} \times \mathcal{V}$,

[2] That $(\mathbb{C}(\Theta, \mathcal{X}), \|\cdot\|_\Theta)$ is a separable Banach space under compact Θ follows from application of the Arzelá-Ascoli theorem to obtain completeness and the Stone-Weierstrass theorem for separability.

let $\bar{x}_\Theta(\cdot) = \phi(\bar{x}, \bar{v}, \cdot) \in \mathbb{C}(\Theta)$. By compactness of Θ and continuity of $\bar{x}_\Theta$ we immediately have $\bar{\bar{x}}_\Theta := \|\bar{x}_\Theta(\cdot)\|_n^\Theta < \infty$. Also $\bar{\bar{\phi}} := \sup_t \|\phi(\bar{x}, v_t, \cdot)\|_n^\Theta < \infty$ by condition *(iii.a)*. Using condition *(iv.a)*, we now have

$$\begin{aligned}
&\sup_t \|x_t(v^{1:t-1}, \cdot, \bar{x}) - \bar{x}_\Theta(\cdot)\|_n^\Theta \\
&\qquad \leq \sup_t \|\phi^*(x_{t-1}(v^{1:t-2}, \cdot, \bar{x}), v_t, \cdot) - \phi(\bar{x}, v_t, \cdot)\|_n^\Theta \\
&\qquad\qquad + \sup_t \|\phi(\bar{x}, v_t, \cdot)\|_n^\Theta + \sup_t \|\bar{x}_\Theta(\cdot)\|_n^\Theta \\
&\qquad \leq \sup_t \|\phi^*(x_{t-1}(v^{1:t-2}, \cdot, \bar{x}), v_t, \cdot) - \phi(\bar{x}, v_t, \cdot)\|_n^\Theta + \bar{\bar{\phi}} + \bar{\bar{x}}_\Theta \\
&\qquad \leq \bar{c} \cdot \sup_t \|x_{t-1}(v^{1:t-2}, \cdot, \bar{x}) - \bar{x}\|_n^\Theta + \bar{\bar{\phi}} + \bar{\bar{x}}_\Theta \\
&\qquad \leq \bar{c} \cdot \sup_t \|x_{t-1}(\cdot, v^{1:t-2}, \bar{x}) - \bar{x}\|_n^\Theta + \bar{\bar{\phi}} + \bar{c}\,\bar{x} + (1+\bar{c})\,\bar{\bar{x}}_\Theta \\
&\qquad \leq \left(\bar{\bar{\phi}} + \bar{c}\,\bar{x} + (1+\bar{c})\,\bar{\bar{x}}_\Theta\right) \sum_{j=1}^{t-2} \bar{c}^j + \sup_t \|\bar{x} - \bar{x}_\Theta(\cdot)\|_n^\Theta \\
&\qquad \leq \frac{\left(\bar{\bar{\phi}} + \bar{c}\,\bar{x} + (1+\bar{c})\,\bar{\bar{x}}_\Theta\right)}{1-\bar{c}} + |\bar{x}| + \bar{\bar{x}}_\Theta < \infty,
\end{aligned}$$

where $\bar{c} < 1$ follows by condition *(iv.a)*.

To use conditions *(iii.b)–(v.b)* instead of conditions *(iii.a)–(v.a)*, we first (re)define $\bar{c}$ as $\bar{c} = (\sup_{\boldsymbol{\theta}\in\Theta} r_t^1(\boldsymbol{\theta}))^{1/n}$. We have

$$\begin{aligned}
&\sup_t \|x_t(v^{1:t-1}, \cdot, \bar{x}) - \bar{x}_\Theta(\cdot)\|_n^\Theta = \\
&\quad \leq \sup_t \|\phi^*(x_{t-1}(v^{1:t-2}, \cdot, \bar{x}), v_t, \cdot) - \phi(\bar{x}, v_t, \cdot)\|_n^\Theta + \bar{\bar{\phi}} + \bar{\bar{x}}_\Theta \\
&\quad \leq \bar{\bar{\phi}} + \bar{\bar{x}}_\Theta + \sup_t \Big(\mathbb{E} \sup_{\boldsymbol{\theta}\in\Theta} |x_{t-1}(v^{1:t-2}, \boldsymbol{\theta}, \bar{x}) - \bar{x}|^n \\
&\qquad\quad \times \sup_{\boldsymbol{\theta}\in\Theta} \frac{|\phi^*(x_{t-1}(v^{1:t-2}, \boldsymbol{\theta}, \bar{x}), v^{1:t-2}, \bar{x}), v_t, \boldsymbol{\theta}) - \phi(\bar{x}, v_t, \boldsymbol{\theta})|^n}{|x_{t-1}(v^{1:t-2}, \boldsymbol{\theta}, \bar{x}) - \bar{x}|^n}\Big)^{1/n} \\
&\quad \leq \bar{\bar{\phi}} + \bar{\bar{x}}_\Theta + \sup_t \Big(\mathbb{E} \sup_{\boldsymbol{\theta}\in\Theta} |x_{t-1}(v^{1:t-2}, \boldsymbol{\theta}, \bar{x}) - \bar{x}|^n \\
&\qquad\quad \times \sup_{\boldsymbol{\theta}\in\Theta} \sup_{(x,x')\in\mathcal{X}\times\mathcal{X}:x\neq x'} \frac{|\phi^*(x, v^{1:t-2}, \bar{x}), v_t, \boldsymbol{\theta}) - \phi(x', v_t, \boldsymbol{\theta})|^n}{|x - x'|^n}\Big)^{1/n} \\
&\quad \overset{\text{by } (v.b)}{\leq} \bar{\bar{\phi}} + (1+\bar{c})\,\bar{\bar{x}}_\Theta + \bar{c}\,\bar{x} + \bar{c} \cdot \sup_t \|x_{t-1}(v^{1:t-2}, \cdot, \bar{x}) - \bar{x}_\Theta(\cdot)\|_n^\Theta.
\end{aligned}$$

Hence, unfolding the process backward in time yields $\sup_t \|x_t(v^{1:t-1}, \cdot, \bar{x}) - \bar{x}_\Theta(\cdot)\|_n^\Theta < \infty$ by the same argument as above.

Finally, using conditions *(iii.c)* and *(iv.c)* instead, we have

$$\begin{aligned}
&\sup_t \|x_t(v^{1:t-1}, \cdot, \bar{x})\|_n^\Theta \leq \sup_t \|\sup_{v\in\mathcal{V}} |\phi^*(x_{t-1}(v^{1:t-2}, \cdot, \bar{x}), v, \cdot)\,|\,\|_n^\Theta \\
&\leq \sup_t \left\|\bar{\phi}\big(x_{t-1}(v^{1:t-2}, \cdot, \bar{x}), \cdot\big) - \bar{\phi}\big(\bar{x}, \cdot\big)\right\|_n^\Theta + \|\bar{\phi}\big(\bar{x}, \cdot\big)\|_n^\Theta \\
&\leq \bar{c} \cdot \sup_t \|x_{t-1}(v^{1:t-2}, \cdot, \bar{x})\|_n^\Theta + \bar{c}\,\bar{x} + \|\bar{\phi}\big(\bar{x}, \cdot\big)\|_n^\Theta
\end{aligned}$$

with $\|\bar{\phi}(\bar{x}, \cdot)\|_n^\Theta < \infty$ by *(iii.c)* and $\bar{c} < 1$ by condition *(iv.c)*. As a result, unfolding the recursion establishes $\sup_t \|x_t(v^{1:t-1}, \cdot, \bar{x})\|_n^\Theta < \infty$ by the same argument as above.

For $0 < n < 1$ the function $\|\cdot\|_n$ is only a pseudo-norm as it is not sub-additive. However, the proof still follows by working instead with the metric $\|\cdot\|_n^* := (\|\cdot\|_n)^n$ which is sub-additive. □

Proof of Proposition 2. The results for the sequence $\{f_t\}$ are obtained by application of Proposition SA.2 with $v_t = y_t$ and $x_t(v^{1:t-1}, \boldsymbol{\theta}, \bar{x}) = f_t(y^{1:t-1}, \boldsymbol{\theta}, \bar{f})$ and $\phi(x_t, v_t, \boldsymbol{\theta}) = \omega + \alpha s(f_t, y_t; \lambda) + \beta f_t$.

Step 1, SE for f_t: Given the compactness of Θ, condition *(i)* directly implies condition *(i)* in Proposition SA.2.

$$\begin{aligned}
\mathbb{E}\log^+ \sup_{\boldsymbol{\theta}\in\Theta} |\phi(\bar{x}, v_t, \boldsymbol{\theta}) - \bar{x}| &= \mathbb{E}\log^+ \sup_{\boldsymbol{\theta}\in\Theta} |\omega + \alpha s(\bar{f}, y_t; \lambda) + \beta\bar{f} - \bar{f}| \\
&\leq \mathbb{E}\log^+ \sup_{\boldsymbol{\theta}\in\Theta} \Big[|\omega| + |\alpha||s(\bar{f}, y_t; \lambda)| + |\beta|\bar{f}| - |\bar{f}|\Big] \\
&\leq \log^+ \sup_{\omega\in\Omega} |\omega| + \log^+ \sup_{\alpha\in\mathcal{A}} |\alpha| + \mathbb{E}\log^+ \sup_{\lambda\in\Lambda} |s(\bar{f}, y_t; \lambda)| \\
&\quad + \sup_{\beta\in\mathcal{B}} \log^+ |(\beta - 1)| + \log^+ |\bar{f}| < \infty
\end{aligned}$$

with $\log^+ \sup_{\omega\in\Omega} | < \infty$, $\log^+ \sup_{\alpha\in\mathcal{A}} |\alpha| < \infty$ and $\sup_{\beta\in\mathcal{B}} \log^+ |(\beta - 1)| < \infty$ by compactness of Θ, and $\log^+ |\bar{f}| < \infty$ for any $\bar{f} \in \mathcal{F} \subseteq \mathbb{R}$ and $\mathbb{E}\log^+ \sup_{\lambda\in\Lambda} |s(\bar{f}, y_t; \lambda)| < \infty$ by condition *(i)* in Proposition 2. Also, condition *(ii)* implies condition *(ii)* in Proposition SA.2 because

$$\begin{aligned}
&\mathbb{E}\log \sup_{\boldsymbol{\theta}\in\Theta} r_1^1(\boldsymbol{\theta}) = \\
&\mathbb{E}\log \sup_{\boldsymbol{\theta}\in\Theta} \sup_{(f,f')\in\mathcal{F}\times\mathcal{F}: f\neq f'} \frac{|\omega - \omega + \alpha(s(f, y_t; \lambda) - s(f', y_t; \lambda)) + \beta(f - f')|}{|f - f'|} \\
&\leq \mathbb{E}\log \sup_{\boldsymbol{\theta}\in\Theta} \sup_{(f,f')\in\mathcal{F}\times\mathcal{F}: f\neq f'} \frac{|\alpha(s(f, y_t; \lambda) - s(f', y_t; \lambda)) + \beta(f - f')|}{|f - f'|} \\
&= \mathbb{E}\log \sup_{\boldsymbol{\theta}\in\Theta} \sup_{(f,f')\in\mathcal{F}\times\mathcal{F}: f\neq f'} \frac{|\alpha\dot{s}_{y,t}(f^*; \lambda)(f - f') + \beta(f - f')|}{|f - f'|} \\
&= \mathbb{E}\log \sup_{\boldsymbol{\theta}\in\Theta} \sup_{f^*\in\mathcal{F}} \Big|\alpha\dot{s}_{y,t}(f^*; \lambda) + \beta\Big| = \mathbb{E}\log \sup_{\boldsymbol{\theta}\in\Theta} \rho_1^1(\boldsymbol{\theta}) < 0.
\end{aligned}$$

Step 2, moment bounds for f_t: By a very similar argument as in Step 1, we can show that condition *(iv)* implies condition *(iv.a)* in Proposition SA.2. Condition *(iii)* implies condition *(iii.b)* in Proposition SA.2 for $n = n_f$ since by the C_r-inequality in (Loève, 1977, p.157), there exists a $0 < c < \infty$ such that

$$\|\phi(\bar{x}, v_t, \cdot)\|_{n_f}^{\Theta} = \mathbb{E} \sup_{\boldsymbol{\theta}\in\Theta} |\omega + \alpha s(\bar{f}, y_t; \lambda) + \beta \bar{f}|^{n_f}$$
$$\leq c \cdot \sup_{\boldsymbol{\theta}\in\Theta} |\omega + \beta \bar{f}|^{n_f} + c \cdot |\alpha|^{n_f} \, \mathbb{E} \sup_{\boldsymbol{\theta}\in\Theta} |s(\bar{f}, y_t; \lambda)|^{n_f} < \infty.$$

Finally, condition *(v)* directly implies condition *(v.b)* in Proposition SA.2. Note also the pointwise independence implies the independence of the suprema since, by continuity and compactness, the supremum is obtained at a point $\boldsymbol{\theta} \in \Theta$.

Step 3, SE for derivatives of f_t: The vector derivative processes initialized at $t = 1$, $\{\boldsymbol{f}_t^{(i)}(y^{1:t-1}, \boldsymbol{\theta}, \bar{\boldsymbol{f}}^{0:i})\}_{t\in\mathbb{N}}$ also satisfies, element by element, the conditions of Theorem 3. For convenience, we now omit the initialization vector, and adopt the notation $f_{i,\boldsymbol{\theta},t} := \boldsymbol{f}_t^{(i)}(y^{1:t-1}, \boldsymbol{\theta}, \bar{\boldsymbol{f}}^{0:i})$ and $\boldsymbol{f}_{0:i,\boldsymbol{\theta},t} := (f_{0,\boldsymbol{\theta},t}, \ldots, f_{i,\boldsymbol{\theta},t})$. Define $\mathbf{v}_{i,t} = (y_t, \boldsymbol{f}_{0:i,\boldsymbol{\theta},t})$. In Appendix B.2 in the technical appendix we show that the dynamic equations generating each element of the vector or matrix partial derivative processes take the form

$$f_{i,\boldsymbol{\theta},t+1} = \boldsymbol{A}_t^{(i)}(\boldsymbol{\theta}) + \boldsymbol{B}_t(\boldsymbol{\theta}) f_{i,\boldsymbol{\theta},t} \quad \forall\ (i, \boldsymbol{\theta}, t), \tag{D.3}$$

with $\boldsymbol{B}_t(\boldsymbol{\theta}) = \boldsymbol{B}_t(\mathbf{v}_{0,t}, \boldsymbol{\theta}) := \beta + \alpha\, \partial s(f_{0,\boldsymbol{\theta},t}, y_t; \lambda)/\partial f \ \forall\ i$, and $\boldsymbol{A}_t^{(i)}(\boldsymbol{\theta}) = \boldsymbol{A}_t^{(i)}(\mathbf{v}_{i-1,t}, \boldsymbol{\theta})$. Note that the 'autoregressive' parameter $\boldsymbol{B}_t(\boldsymbol{\theta})$ does not depend on i and is therefore the same for all derivative processes. Furthermore, by the expression in Section B.2 of the technical appendix, we have

$$\mathbb{E} \log^+ \sup_{\boldsymbol{\theta}\in\Theta} |\boldsymbol{A}_t^{(i)}(\boldsymbol{\theta})|^{n_f^{(i)}} \leq \mathbb{E} \sup_{\boldsymbol{\theta}\in\Theta} |\boldsymbol{A}_t^{(i)}(\boldsymbol{\theta})|^{n_f^{(i)}} < \infty, \tag{D.4}$$

for $i = 1, 2$. Conditions *(i)* and *(ii)* now directly imply conditions *(i)* and *(ii)* in Proposition SA.2 for both the first and second derivative processes. Since all derivative sequences follow (D.3) and $s \in \mathbb{C}^{(2,0,2)}(\mathcal{Y} \times \mathcal{F} \times \Theta)$, this implies that $\{\boldsymbol{A}_t^{(i)}(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ and $\{\boldsymbol{B}_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ are continuous functions of $\{y_t\}_{t\in\mathbb{Z}}$ and $\{\boldsymbol{f}_t^{(0:i-1)}(y^{1:t-1}, \boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ and hence SE. As a result, since $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |\boldsymbol{B}_t(\boldsymbol{\theta})| < 1$, it follows that $\sup_{\boldsymbol{\theta}\in\Theta} \|\boldsymbol{f}_t^{(i)}(\boldsymbol{\theta}, \bar{\boldsymbol{f}}^{0:i}) - \boldsymbol{f}_t^{(i)}(\boldsymbol{\theta})\| \stackrel{e.a.s.}{\to} 0$ as $t \to \infty$, $i = 0, 1, 2$.

Step 4, moment bounds for derivatives of f_t: To establish the existence of moments for the derivative processes, we also need to verify that conditions *(iii.b)*–*(v.b)* in Proposition SA.2 hold for these processes.

Condition *(iii)* implies condition *(iii.b)* in Proposition SA.2 for derivative process i with $n = n_f^{(i)}$ for $i = 1, 2$: from the C_r-inequality in (Loève, 1977, p.157), there exists a $0 < c < \infty$ such that,

$$\begin{aligned}\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|\phi(\bar{x}, v_t, \boldsymbol{\theta})|^{n_f^{(i)}} = \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|\boldsymbol{A}_t^{(i)}(\boldsymbol{\theta}) + \boldsymbol{B}_t(\boldsymbol{\theta})\bar{\boldsymbol{f}}^{(i)}|^{n_f^{(i)}} \quad &\leq \\ c\cdot\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|\boldsymbol{A}_t^{(i)}(\boldsymbol{\theta})|^{n_f^{(i)}} + c\cdot|\bar{\boldsymbol{f}}^{(i)}|^{n_f^{(i)}}\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|\boldsymbol{B}_t(\boldsymbol{\theta})|^{n_f^{(i)}} < \infty&,\end{aligned}$$

using (D.4) and $n_f^{(2)} \leq n_f^{(1)} \leq n_f$.

Condition *(iv.b)* in Proposition SA.2 follows by noting that for $i = 1, 2$ we have

$$\begin{aligned}&\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} r_1^{n_f^{(i)}}(\boldsymbol{\theta}) \leq \\ &\leq \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\sup_{(f,f')\in\mathcal{F}\times\mathcal{F}:f\neq f'}\frac{|\alpha(s(f, y_t; \lambda) - s(f', y_t; \lambda)) + \beta(f - f')|^{n_f^{(i)}}}{|f - f'|^{n_f^{(i)}}} \\ &= \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\sup_{(f,f')\in\mathcal{F}\times\mathcal{F}:f\neq f'}\frac{|\alpha\dot{s}_{y,t}(f^*; \lambda)(f - f') + \beta(f - f')|^{n_f^{(i)}}}{|f - f'|^{n_f^{(i)}}} \\ &= \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\sup_{f^*\in\mathcal{F}}\left|\alpha\dot{s}_{y,t}(f^*; \lambda) + \beta\right|^{n_f^{(i)}} = \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\rho_1^{n_f^{(i)}}(\boldsymbol{\theta}) < 1.\end{aligned}$$

Finally, condition *(v.b)* directly implies condition *(v.b)* in Proposition SA.2 by continuity of the derivative operator and the fact that continuous transformations of independent random variables are independent. □

E Technical Lemmas and Proofs

Lemma SA.1 provides a proof for the alternative bound in Remark 2 of the paper.

Lemma SA.1. *Condition (ii) in Proposition 1 can be substituted by the condition*

$$\sum_{k=0}^{n_f}\binom{n_f}{k}|\alpha|^k\,|\beta|^{n_f-k}\,\mathbb{E}\sup_{f^*\in\mathcal{F}^*}|\dot{s}_{u,t+1}(f^*; \lambda)|^k < 1.$$

Proof. Noting that for $k = n \geq 1$ the Binomial theorem implies

$$\begin{aligned}
\mathbb{E} \sup_{f \in \mathcal{F}} \left| \alpha \frac{\partial s_u(u_t, f; \lambda)}{\partial f} + \beta \right|^n &\leq \mathbb{E} \sup_{f^* \in \mathcal{F}^*} \left(|\alpha| \left| \frac{\partial s_u(u_t, f; \lambda)}{\partial f} \right| + |\beta| \right)^n \\
&= \mathbb{E} \sup_{f^* \in \mathcal{F}^*} \sum_{k=0}^{n} \binom{n}{k} |\alpha|^k \left| \frac{\partial s_u(u_t, f; \lambda)}{\partial f} \right|^k |\beta|^{n-k} \\
&\leq \sum_{k=0}^{n} \binom{n}{k} |\alpha|^k |\beta|^{n-k} \mathbb{E} \sup_{f^* \in \mathcal{F}^*} \left| \frac{\partial s_u(u_t, f; \lambda)}{\partial f} \right|^k .
\end{aligned}$$

□

Lemma SA.2 provides a range of alternative bounds that are applicable depending on the properties of the score. This lemma is useful in applications for deriving bounds on the parameter. The lemma uses a bound on the score process s_u that is splittable in a function $\eta(f;\lambda)$ that depends on f, and a function $\zeta(v;\lambda)$ that depends on v.

Lemma SA.2. *Let $\bar{s}'_{u,k}(\lambda) \in \mathbb{R}_0^+$ be a constant $\forall\ \lambda \in \Lambda$ satisfying the bound*

$$\mathbb{E} \sup_{f} \left| s_u(f, v; \lambda) \right|^k < \bar{s}'_{u,k}(\lambda)\ \forall\ 1 \leq k \leq n_f,$$

where $s_u(f, v; \lambda) := \partial s_u(f, v, \lambda)/\partial f$.

(a) If $|s_u(f, v; \lambda)| \leq |\eta(f;\lambda)\zeta(v;\lambda)|\ \forall\ (f, v, \lambda)$, then we can set

$$\bar{s}'_{u,k}(\lambda) = \sup_{f} |\eta(f;\lambda)|^k \mathbb{E} |\zeta(v_t(\lambda);\lambda)|^k.$$

(b) If $|s_u(f, v; \lambda)| \leq \sum_{j=1}^{J} |\eta_j(f;\lambda)\zeta_j(v;\lambda)|\ \forall\ (f, v, \lambda)$, then we can set

$$\bar{s}'_{u,k}(\lambda) = \sum_{j_1 + \ldots + j_J = k} \binom{k}{j_1, \ldots, j_J} \prod_{\iota=1}^{J} \sup_{f} |\eta_\iota(f;\lambda)|^{j_\iota} \mathbb{E} \left| \prod_{\iota=1}^{J} \zeta_\iota(v_t(\lambda);\lambda) \right|^{j_\iota}.$$

(c) If $|\eta_\iota(f;\lambda)| \leq |\eta(f;\lambda)|\ \forall\ (f, \lambda, \iota) \in \mathcal{F} \times \Lambda \times \{1, \ldots, J\}$ and $|\zeta_\iota(v;\lambda)| \leq |\zeta(v;\lambda)|\ \forall\ (v, \lambda, \iota) \in \mathcal{V} \times \Lambda \times \{1, \ldots, J\}$, then we can set

$$\bar{s}'_{u,k}(\lambda) = J^k \sup_{f} |\eta(f;\lambda)|^k \mathbb{E} |\zeta(v_t;\lambda)|^k.$$

(d) If $s_u(f, v; \lambda) = \sum_{j=1}^{J} \eta_j^*(f;\lambda)\zeta_j^*(v;\lambda)\ \forall\ (f, v, \lambda)$. Then, (b) and (c) hold with $\eta_j(f;\lambda) = \partial \eta_j^*(f;\lambda)/\partial f$ and $\zeta_j(v;\lambda) = \zeta_j^*(v;\lambda)\ \forall\ j$.

Proof. Part *(b)* follows by applying the multinomial theorem and noting that

$$\begin{aligned}
&\mathbb{E}\sup_f \Big|\frac{\partial s_u(v_t(\lambda), f; \lambda)}{\partial f}\Big|^k \\
&\leq \mathbb{E}\sup_f \Big|\sum_{j=1}^J \eta_j(f;\lambda)\zeta_j(v_t(\lambda);\lambda)\Big|^k \leq \mathbb{E}\sup_f \Big(\sum_{j=1}^J |\eta_j(f;\lambda)||\zeta_j(v_t(\lambda);\lambda)|\Big)^k \\
&\leq \mathbb{E}\sup_f \sum_{j_1+\ldots+j_J=k} \binom{k}{j_1,\ldots,j_J} \prod_{\iota=1}^J |\eta_\iota(f;\lambda)|^{j_\iota} |\zeta_\iota(v_t(\lambda);\lambda)|^{j_\iota} \\
&\leq \sum_{j_1+\ldots+j_J=k} \binom{k}{j_1,\ldots,j_J} \sup_f \prod_{\iota=1}^J |\eta_\iota(f;\lambda)|^{j_\iota} \mathbb{E} \prod_{\iota=1}^J |\zeta_\iota(v_t(\lambda);\lambda)|^{j_\iota} \\
&= \sum_{j_1+\ldots+j_J=k} \binom{k}{j_1,\ldots,j_J} \prod_{\iota=1}^J \sup_f |\eta_\iota(f;\lambda)|^{j_\iota} \mathbb{E}\Big| \prod_{\iota=1}^J \zeta_\iota(v_t(\lambda);\lambda)\Big|^{j_\iota}
\end{aligned}$$

where we use the multinomial coefficient

$$\binom{k}{j_1,\ldots,j_J} = \frac{k!}{j_1!\cdots j_J!}.$$

Part *(a)* now follows immediately from *(b)* under $J = 1$. Also part *(c)* follows from *(b)*, since

$$\begin{aligned}
&\mathbb{E}\sup_f \Big|\frac{\partial s_u(v_t(\lambda), f; \lambda)}{\partial f}\Big|^k \\
&\leq \sum_{j_1+\ldots+j_J=k} \binom{k}{j_1,\ldots,j_J} \prod_{\iota=1}^J \sup_f |\eta_\iota(f;\lambda)|^{j_\iota} \mathbb{E}\Big| \prod_{\iota=1}^J \zeta_\iota(v_t(\lambda);\lambda)\Big|^{j_\iota}. \\
&= \sum_{j_1+\ldots+j_J=k} \binom{k}{j_1,\ldots,j_J} \sup_f \prod_{\iota=1}^J |\eta_\iota(f;\lambda)|^{j_\iota} \mathbb{E} \prod_{\iota=1}^J \Big|\zeta_\iota(v_t(\lambda);\lambda)\Big|^{j_\iota}.
\end{aligned}$$

Now, under the maintained assumptions,

$$\begin{aligned}
&\sum_{j_1+\ldots+j_J=k} \binom{k}{j_1,\ldots,j_J} \sup_f \prod_{\iota=1}^J |\eta_\iota(f;\lambda)|^{j_\iota} \mathbb{E} \prod_{\iota=1}^J \Big|\zeta_\iota(v_t(\lambda);\lambda)\Big|^{j_\iota} \\
&\qquad \leq \sum_{j_1+\ldots+j_J=k} \binom{k}{j_1,\ldots,j_J} \sup_f \prod_{\iota=1}^J |\eta(f;\lambda)|^{j_\iota} \mathbb{E} \prod_{\iota=1}^J \Big|\zeta(v_t(\lambda);\lambda)\Big|^{j_\iota} \\
&\qquad \leq \sum_{j_1+\ldots+j_J=k} \binom{k}{j_1,\ldots,j_J} \sup_f |\eta(f;\lambda)|^k \mathbb{E}|\zeta(v_t(\lambda);\lambda)|^k \\
&\qquad \leq \sum_{j_1+\ldots+j_J=k} \binom{k}{j_1,\ldots,j_J} \bar{\eta}_k(\lambda)\bar{\zeta}_k(\lambda) = J^k \bar{\eta}_k(\lambda)\bar{\zeta}_k(\lambda),
\end{aligned}$$

because $\sum_{j_1+\ldots+j_J=k} \binom{k}{j_1,\ldots,j_J} = J^k$.

The proof of (d) follows from *(b)* by the fact that under the maintained assumptions

$$\frac{\partial s_u(f,v;\lambda)}{\partial f} = \sum_{j=1}^{J} \frac{\partial \eta_j^*(f;\lambda)}{\partial f} \zeta_j^*(v;\lambda) \ \forall \ (f,v;\lambda).$$

$\mathbb{E}|s_u(v_t(\lambda), f^*;\lambda)|^k < \infty$ is satisfied since,

$$\big(\mathbb{E}|s_u(v_t(\lambda), f^*;\lambda)|^k\big)^{\frac{1}{k}} \leq \sum_{j=1}^{J} |\eta_j^*(f^*;\lambda)| \big(\mathbb{E}|\zeta_j^*(v;\lambda)|^k\big)^{\frac{1}{k}} < \infty.$$

Finally, $s \in \mathbb{M}_{\boldsymbol{\theta},\boldsymbol{\theta}}(n, m^{(\boldsymbol{k})})$ because it is a sum of products of derivatives of η_j^* w.r.t. f and λ and derivatives of ζ_j^* w.r.t. λ, which are all $\mathbb{M}_{\boldsymbol{\theta},\boldsymbol{\theta}}(n, m^{(\boldsymbol{k})})$ because η_j^* is uniformly bounded and ζ_j^* is $\mathbb{M}_{\boldsymbol{\theta},\boldsymbol{\theta}}(n_v, m^{(\boldsymbol{k})})$. □

The following set of lemmas derives the bounds on the moments of the likelihood function and its derivatives given the moment preserving maps assumed in the main text.

Lemma SA.3. *Let* $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |f_t(\boldsymbol{\theta})|^{n_f} < \infty$*,* $\mathbb{E}|y_t|^{n_y} < \infty$*,* $\boldsymbol{n} = (n_f, n_y)$*,* $\tilde{g}' \in \mathbb{M}(\boldsymbol{n}, n_{\log \tilde{g}'})$ *and* $\tilde{p} \in \mathbb{M}(\boldsymbol{n}, n_{\tilde{p}})$*. Then* $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |\ell_T(\boldsymbol{\theta}, f)|^m < \infty$ *where* $m = \min\{n_{\log \tilde{g}'}, n_{\tilde{p}}\}$*.*

Proof. The statement follows immediately from the fact that

$$\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |\ell_T(\boldsymbol{\theta}, f)| \leq (1/T) \sum_{t=1}^{T} \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |\tilde{p}_t| + \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |\log \tilde{g}_t'|.$$

□

Lemma SA.4. *Let* $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |f_t(\boldsymbol{\theta})|^{n_f} < \infty$*,* $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |\boldsymbol{f}_t^{(1)}(\boldsymbol{\theta})|^{n_f^{(1)}} < \infty$ *and* $\mathbb{E}|y_t|^{n_y} < \infty$*, and suppose that, for* $\boldsymbol{n} := (n_f, n_y)$*, it holds that*

$$\begin{aligned} s^{(\boldsymbol{k})} &\in \mathbb{M}_{\Theta,\Theta}(\boldsymbol{n}, n_s^{(\boldsymbol{k})}), \qquad \tilde{p}^{(\boldsymbol{k}')} \in \mathbb{M}_{\Theta,\Theta}(n_{\tilde{g}}, n_{\tilde{p}}^{(\boldsymbol{k}')}), \\ (\log \tilde{g}')^{(\boldsymbol{k}'')} &\in \mathbb{M}_{\Theta,\Theta}(\boldsymbol{n}, n_{\log \tilde{g}'}^{(\boldsymbol{k}'')}). \end{aligned}$$

Then $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} |\ell_T'(\boldsymbol{\theta}, f)|^m < \infty$ *where*

$$m = \min\left\{ n_{\tilde{p}}^{(0,0,1)} \ , \ \frac{n_{\log \tilde{g}'}^{(1,0)} n_f^{(1)}}{n_{\log \tilde{g}'}^{(1,0)} + n_f^{(1)}} \ , \ \frac{n_{\tilde{p}}^{(1,0,0)} n_f^{(1)}}{n_{\tilde{p}}^{(1,0,0)} + n_f^{(1)}} \right\}. \tag{E.1}$$

Proof. The statement follows by Holder's generalized inequality and the explicit form of the first derivative of the likelihood in (B.3) in Appendix B, and the properties for moment preserving maps in Lemma SA.7. □

Lemma SA.5. *Let* $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|f_t(\boldsymbol{\theta})|^{n_f}<\infty$, $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|\boldsymbol{f}_t^{(1)}(\boldsymbol{\theta})|^{n_f^{(1)}}<\infty$, $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|\boldsymbol{f}_t^{(2)}(\boldsymbol{\theta})|^{n_f^{(2)}}<\infty$ *and* $\mathbb{E}|y_t|^{n_y}<\infty$, *For* $\boldsymbol{n}:=(n_f,n_y)$ *it holds that*

$$\begin{aligned} s^{(\boldsymbol{k})} &\in \mathbb{M}_{\Theta,\Theta}(\boldsymbol{n},n_s^{(\boldsymbol{k})}), \qquad \tilde{p}^{(\boldsymbol{k}')}\in\mathbb{M}_{\Theta,\Theta}(n_{\tilde{g}},n_{\tilde{p}}^{(\boldsymbol{k}')}),\\ (\log\tilde{g}')^{(\boldsymbol{k}'')} &\in \mathbb{M}_{\Theta,\Theta}(\boldsymbol{n},n_{\log\tilde{g}'}^{(\boldsymbol{k}'')}). \end{aligned}$$

Then $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|\ell_T''(\boldsymbol{\theta},f)|^m<\infty$ *where*

$$\begin{aligned} m &= \min\Bigg\{n_{\tilde{p}}^{(0,0,2)}\ ,\ \frac{n_{\tilde{p}}^{(1,0,1)}n_f^{(1)}}{n_{\tilde{p}}^{(1,0,1)}+n_f^{(1)}}\ ,\ \frac{n_{\tilde{p}}^{(2,0,0)}n_f^{(1)}}{2n_{\tilde{p}}^{(2,0,0)}+n_f^{(1)}}\ ,\\ &\qquad \frac{n_{\tilde{p}}^{(1,0,0)}n_f^{(2)}}{n_{\tilde{p}}^{(1,0,0)}+n_f^{(2)}}\ ,\ \frac{n_{\log\tilde{g}'}^{(1,0)}n_f^{(2)}}{n_{\log\tilde{g}'}^{(1,0)}+n_f^{(2)}}\ ,\ \frac{n_{\log\tilde{g}'}^{(2,0)}n_f^{(1)}}{2n_{\log\tilde{g}'}^{(2,0)}+n_f^{(1)}}\Bigg\}. \qquad \text{(E.2)} \end{aligned}$$

Proof. The statement follows by Holder's generalized inequality and the explicit form of the second derivative of the likelihood in (B.3) in Appendix B, and the properties for moment preserving maps in Lemma SA.7. □

The following lemma shows that $\partial\ell'(y^{1:t},\hat{\boldsymbol{f}}_t^{(0:1)})/\partial f$ is bounded in probability under the assumptions maintained in Theorem 4.

Lemma SA.6. *Let the conditions of Theorem 4 hold true. Then*

$$\big|\partial\ell'(y^{1:t},\hat{\boldsymbol{f}}_t^{(0:1)})/\partial f\big| = O_p(1) \text{ in } t.$$

Proof. $\big|\partial\ell'(y^{1:t},\hat{\boldsymbol{f}}_t^{(0:1)})/\partial f\big| = O_p(1)$ is obtained from $\mathbb{E}\big|\frac{\partial\ell'(y^{1:t},\hat{\boldsymbol{f}}_t^{(0:1)})}{\partial f}\big|<\infty$ which is implied by $\mathbb{E}|y_t|^{n_y}<\infty$, $\mathbb{E}|\hat{f}_t|^{n_f}<\infty$, $\mathbb{E}\|\hat{f}_t^{(1)}\|^{n_f^{(1)}}<\infty$ and the fact that, inspection of the likelihood expressions in the technical appendix reveals that, $\partial\ell'/\partial f_1\in\mathbb{M}(\boldsymbol{n},n_{\ell'}^f)$ with $\boldsymbol{n}=(n_f,n_f^{(1)},n_y)$ and

$$n_{\ell'}^f := \min\Bigg\{\frac{n_{\tilde{p}}^{(2,0,0)}n_f^{(1)}}{n_{\tilde{p}}^{(2,0,0)}+n_f^{(1)}}\ ,\ \frac{n_{\log\tilde{g}'}^{(2,0)}n_f^{(1)}}{n_{\log\tilde{g}'}^{(2,0)}+n_f^{(1)}}\Bigg\}$$

which satisfies $n_{\ell'}^f\geq 1$ by Assumption 7, and $\partial\ell'/\partial f_j\in\mathbb{M}_{\Theta,\Theta}(\boldsymbol{n},n_{\ell'}^{f'})$, for $j=2,...,5$, with $n_{\ell'}^f:=\min\{n_{\tilde{p}}^{(1,0,0)}\ ,\ n_{\log\tilde{g}'}^{(1,0)}\}$ which also satisfies $n_{\ell'}^{f'}\geq 1$ by Assumption 7. Note also that $\mathbb{E}|\hat{f}_t|^{n_f}<\infty$ and $\mathbb{E}\|\hat{f}_t^{(1)}\|^{n_f^{(1)}}<\infty$ follows from the fact that $\hat{\boldsymbol{f}}_t^{(0:1)}$ is a point between $\boldsymbol{f}_t^{(0:1)}(y^{1:t-1},\boldsymbol{\theta}_0,\boldsymbol{f}_1^{(0:1)})$ and $\boldsymbol{f}_t^{(0:1)}(y^{1:t-1},\boldsymbol{\theta}_0)$ for every t, where both bounds satisfy the desired moment condition uniformly in t by Proposition SA.2 under the maintained assumptions. □

The final lemma provides simple moment preserving properties for several common functions of random variables. For notational simplicity we let $h\in$

$\mathbb{M}_{\Theta,\Theta}(n,m)$ denote a function whose h^{th} derivative is an element of the set $\mathbb{M}_{\Theta,\Theta}(n,m)$. In other words, we have $h \in \mathbb{M}^{k}_{\Theta,\Theta}(n,m) \Leftrightarrow h^{(k)} \in \mathbb{M}_{\Theta,\Theta}(n,m)$.

Lemma SA.7. (Catalog of $\mathbb{M}^{\boldsymbol{k}}_{\Theta,\Theta}(\boldsymbol{n},m)$ Moment Preserving Maps) *For every $\boldsymbol{\theta} \in \Theta$, let $h(\cdot;\boldsymbol{\theta}) : \mathcal{X} \to \mathbb{R}$ and $w(\cdot,\cdot,\boldsymbol{\theta}) : \mathcal{X} \times \mathcal{V} \to \mathbb{R}$ be measurable functions.*

(a) Let $h(\cdot;\boldsymbol{\theta})$ be an affine function, $h(x;\boldsymbol{\theta}) = \theta_0 + \theta_1 x \ \forall\ (x,\boldsymbol{\theta}) \in \mathcal{X} \times \Theta$, $\boldsymbol{\theta} = (\theta_0,\theta_1) \in \Theta \subseteq \mathbb{R}^2$. Then, $h(\cdot;\boldsymbol{\theta}) \in \mathbb{M}^{k}_{\Theta,\boldsymbol{\theta}}(n,m)$ with $n = m \ \forall\ \boldsymbol{\theta} \in \Theta \wedge k = 0$ and $h(\cdot;\boldsymbol{\theta}) \in \mathbb{M}^{k}_{\Theta,\boldsymbol{\theta}}(n,m) \ \forall\ (\boldsymbol{\theta},n,m,k) \in \Theta \times \mathbb{R}^{+}_{0} \times \mathbb{R}^{+}_{0} \times \mathbb{N}$. If Θ is compact, then $h \in \mathbb{M}^{k}_{\Theta,\Theta}(n,m)$ with $n = m$ for $k = 0$ and $h(\cdot;\boldsymbol{\theta}) \in \mathbb{M}^{k}_{\Theta,\Theta}(n,m) \ \forall\ (n,m,k) \in \mathbb{R}^{+}_{0} \times \mathbb{R}^{+}_{0} \times \mathbb{N}$.

(b) Let $h(\cdot;\boldsymbol{\theta})$ be a polynomial function, $h(x;\boldsymbol{\theta}) = \sum_{j=0}^{J} \theta_j x^j \ \forall\ (x,\boldsymbol{\theta}) \in \mathcal{X} \times \Theta$, $\boldsymbol{\theta} = (\theta_0,\ldots,\theta_J) \in \Theta \subseteq \mathbb{R}^J$, $J \geq 1$. Then $h(\cdot;\boldsymbol{\theta}) \in \mathbb{M}^{k}_{\Theta,\boldsymbol{\theta}}(n,m)$ with $m = n/(J-k) \ \forall\ (k,\boldsymbol{\theta}) \in \mathbb{N}_0 \times \Theta$. If Θ is compact, then $h \in \mathbb{M}^{k}_{\Theta,\Theta}(n,m)$ with $m = n/(J-k) \ \forall\ k \in \mathbb{N}_0$.

(c) Let $h(x;\boldsymbol{\theta}) = \sum_{j=0}^{J} \theta_j x^{r_j} \ \forall\ (x,\boldsymbol{\theta}) \in \mathcal{X} \times \Theta$, $\boldsymbol{\theta} = (\theta_0,\ldots,\theta_J) \in \Theta \subseteq \mathbb{R}^J$ where $r_j \geq 0$. Then $h(\cdot;\boldsymbol{\theta}) \in \mathbb{M}^{k}_{\Theta,\boldsymbol{\theta}}(n,m)$ with $m = n/(\max_j r_j - k) \ \forall\ (\boldsymbol{\theta},k) \in \Theta \in \mathbb{N}_0 : k \leq \min_j r_j$. If Θ is compact, then $h \in \mathbb{M}^{k}_{\Theta,\Theta}(n,m)$ with $m = n/(\max_j r_j - k) \ \forall\ k \in \mathbb{N}_0 : k \leq \min_j r_j$.

(d) Let $\sup_{x\in\mathcal{X}} |h(x;\boldsymbol{\theta})| \leq \bar{h}(\boldsymbol{\theta}) < \infty \ \forall\ \boldsymbol{\theta} \in \Theta$. Then $h(\cdot;\boldsymbol{\theta}) \in \mathbb{M}^{0}_{\Theta,\boldsymbol{\theta}}(n,m)$ $\forall\ (n,m,\boldsymbol{\theta}) \in \Theta \times \mathbb{R}^{+}_{0} \times \mathbb{R}^{+}_{0}$. If additionally, $\sup_{\boldsymbol{\theta}\in\Theta} \bar{h}(\boldsymbol{\theta}) \leq \bar{\bar{h}} < \infty$, then $h \in \mathbb{M}^{0}_{\Theta,\Theta}(n,m) \ \forall\ (n,m) \in \mathbb{R}^{+}_{0} \times \mathbb{R}^{+}_{0}$.

(e) Let $h(\cdot;\boldsymbol{\theta}) \in \mathbb{C}^k(\mathcal{X})$ and $\sup_{x\in\mathcal{X}} |h^{(k)}(x;\boldsymbol{\theta})| \leq \bar{h}_k(\boldsymbol{\theta}) < \infty \ \forall\ \boldsymbol{\theta} \in \Theta$. Then $h(\cdot;\boldsymbol{\theta}) \in \mathbb{M}^{k}_{\Theta,\boldsymbol{\theta}}(n,m)$ with $m = n/k \ \forall\ \boldsymbol{\theta} \in \Theta$. If furthermore, $\sup_{\boldsymbol{\theta}\in\Theta} \bar{h}_k(\boldsymbol{\theta}) \leq \bar{\bar{h}} < \infty$, then $h \in \mathbb{M}^{k}_{\Theta,\Theta}(n,m)$ with $m = n/k$.

(f) Let $w(x,v;\boldsymbol{\theta}) = \theta_0 + \theta_1 x + \theta_2 v$, $(\theta_0,\theta_1,\theta_2,x,v) \in \mathbb{R}^3 \times \mathcal{X} \times \mathcal{V}$. Then $w(\cdot,\cdot,\boldsymbol{\theta}) \in \mathbb{M}^{(k_x,k_v)}_{\Theta,\boldsymbol{\theta}}(\boldsymbol{n},m) \ \forall\ (k_x,k_v,\boldsymbol{\theta}) \in \mathbb{N}_0 \times \mathbb{N}_0 \times \Theta$ with $\boldsymbol{n} = (n_x,n_v)$ and $m = \min\{n_x,n_v\}$. If furthermore Θ is compact, then $w \in \mathbb{M}^{(k_x,k_v)}_{\Theta,\Theta}(\boldsymbol{n},m) \ \forall\ (k_x,k_v) \in \mathbb{N}_0 \times \mathbb{N}_0$ with $m = \min\{n_x,n_v\}$;

(g) If $w(x,v,\boldsymbol{\theta}) = \theta_0 + \theta_1 xv$, $(\theta_0,\theta_1) \in \mathbb{R}^2$ then $w(\cdot,\cdot,\boldsymbol{\theta}) \in \mathbb{M}^{(k_x,k_v)}_{\Theta,\boldsymbol{\theta}}(\boldsymbol{n},m) \ \forall$ $(k_x,k_v,\boldsymbol{\theta}) \in \mathbb{N}_0 \times \mathbb{N}_0 \times \Theta$ with $\boldsymbol{n} = (n_x,n_v)$ where $m = n_x n_v/(n_x+n_v)$. If furthermore, Θ is compact, then $w \in \mathbb{M}^{(k_x,k_v)}_{\Theta,\Theta}(\boldsymbol{n},m) \ \forall\ (k_x,k_v) \in \mathbb{N}_0 \times \mathbb{N}_0$ with $\boldsymbol{n} = (n_x,n_v)$ where $m = n_x n_v/(n_x+n_v)$.

Proof. By the C_r-inequality in (Loève, 1977, p.157), for *(a)* we have, for some c, $\mathbb{E}|h(x_t(\boldsymbol{\theta});\boldsymbol{\theta})|^n = \mathbb{E}|\theta_0 + \theta_1 x_t(\boldsymbol{\theta})|^n \leq c\mathbb{E}|\theta_0|^n + c\mathbb{E}|\theta_1 x_t(\boldsymbol{\theta})|^n \leq c|\theta_0|^n + c|\theta_1|^n\mathbb{E}|x_t(\boldsymbol{\theta})|^n$, and hence, $h(\cdot;\boldsymbol{\theta}) \in \mathbb{M}^0_{\Theta,\boldsymbol{\theta}}(n,m)$ with $n = m \ \forall \ \boldsymbol{\theta} \in \Theta$ because $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|x_t(\boldsymbol{\theta})|^n < \infty \Rightarrow \mathbb{E}|x_t(\boldsymbol{\theta})|^n < \infty \ \forall \ \boldsymbol{\theta} \in \Theta \Rightarrow \mathbb{E}|h(x_t(\boldsymbol{\theta});\boldsymbol{\theta})|^n < \infty \ \forall \ \boldsymbol{\theta} \in \Theta$. Also, $h(\cdot;\boldsymbol{\theta}) \in \mathbb{M}^k_{\Theta,\boldsymbol{\theta}}(n,m) \ \forall \ (m,n,k,\boldsymbol{\theta}) \in \mathbb{R}^+_0 \times \mathbb{R}^+_0 \times \mathbb{N} \times \Theta$ as $h^{(1)}(x_t(\boldsymbol{\theta}),\boldsymbol{\theta}) = \theta_1$ and $h^{(i)}(x_t(\boldsymbol{\theta}),\boldsymbol{\theta}) = 0 \ \forall \ i \geq 2$. Furthermore, $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|h(x_t(\boldsymbol{\theta});\boldsymbol{\theta})|^n = \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|\theta_0+\theta_1 x_t(\boldsymbol{\theta})|^n \leq c\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|\theta_0|^n + c\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|\theta_1 x_t(\boldsymbol{\theta})|^n \leq c\sup_{\boldsymbol{\theta}\in\Theta}|\theta_0|^n + c\sup_{\boldsymbol{\theta}\in\Theta}|\theta_1|^n\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|x_t(\boldsymbol{\theta})|^n$ and as a result, if Θ is compact, we have $h \in \mathbb{M}^0_{\Theta,\Theta}(n,m)$ with $n = m$ because $\sup_{\boldsymbol{\theta}\in\Theta}|\theta_0|^n < \infty$ and $\sup_{\boldsymbol{\theta}\in\Theta}|\theta_1|^n < \infty$, and hence, $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|x_t(\boldsymbol{\theta})|^n < \infty \Rightarrow \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|h\ (x_t(\boldsymbol{\theta});\boldsymbol{\theta})|^n < \infty$. Again, $h \in \mathbb{M}^k_{\Theta,\Theta}(n,m) \ \forall \ (m,n,k) \in \mathbb{R}^+_0 \times \mathbb{R}^+_0 \times \mathbb{N}$ follows from having $h^{(1)}(x_t(\boldsymbol{\theta}),\boldsymbol{\theta}) = \theta_1$ and $h^{(i)}(x_t(\boldsymbol{\theta}),\boldsymbol{\theta}) = 0 \ \forall \ i \geq 2$.

For *(b)* we have that, for some c, $\mathbb{E}|h(x_t(\boldsymbol{\theta});\boldsymbol{\theta})|^n = \mathbb{E}|\sum_{j=0}^J \theta_j x_t^j(\boldsymbol{\theta})|^n \leq c\sum_{j=0}^J \mathbb{E}|\theta_j x_t^j(\boldsymbol{\theta})|^n \leq c\sum_{j=0}^J |\theta_j|^n\mathbb{E}|x_t(\boldsymbol{\theta})|^{jn}$, and hence, $h(\cdot;\boldsymbol{\theta}) \in \mathbb{M}^0_{\Theta,\boldsymbol{\theta}}(n,m)$ with $m = n/J \ \forall \ \boldsymbol{\theta} \in \Theta$ because $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|x_t(\boldsymbol{\theta})|^n < \infty \Rightarrow \mathbb{E}|x_t(\boldsymbol{\theta})|^n < \infty \ \forall \ \boldsymbol{\theta} \in \Theta \Rightarrow \mathbb{E}|h(x_t(\boldsymbol{\theta});\boldsymbol{\theta})|^{n/J} \leq c\sum_{j=0}^J |\theta_j|^n\mathbb{E}|x_t(\boldsymbol{\theta})^j|^{n/J} \leq c \cdot J \cdot \mathbb{E}|x_t(\boldsymbol{\theta})|^n < \infty \ \forall \ \boldsymbol{\theta} \in \Theta$. Also, $h(\cdot;\boldsymbol{\theta}) \in \mathbb{M}^k_{\Theta,\boldsymbol{\theta}}(n,m) \ \forall \ (k,\boldsymbol{\theta}) \in \mathbb{N}_0 \times \Theta$ with $m = n/(J-k)$ because $h^{(k)}(x_t(\boldsymbol{\theta}),\boldsymbol{\theta}) = \sum_{j=k}^J \theta_j^* x^{j-k}$ and hence $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|x_t(\boldsymbol{\theta})|^n < \infty \Rightarrow \mathbb{E}|x_t(\boldsymbol{\theta})|^n < \infty \ \forall \ \boldsymbol{\theta} \in \Theta \Rightarrow \mathbb{E}|h^{(k)}(x_t(\boldsymbol{\theta});\boldsymbol{\theta})|^{n/(J-k)} \leq c\sum_{j=0}^J \mathbb{E}|\theta_j^* x_t(\boldsymbol{\theta})^{j-k}|^{n/(J-k)} \leq c\sum_{j=0}^J |\theta_j^*|^{n/(J-k)}\mathbb{E}|x_t(\boldsymbol{\theta})|^n < \infty \ \forall \ \boldsymbol{\theta} \in \Theta$. Furthermore, $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|h(x_t(\boldsymbol{\theta});\ \boldsymbol{\theta})|^n = \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|\sum_{j=0}^J \theta_j x_t^j(\boldsymbol{\theta})|^n \leq c\sum_{j=0}^J \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|\theta_j x_t^j(\boldsymbol{\theta})|^n \leq c\sum_{j=0}^J \sup_{\boldsymbol{\theta}\in\Theta}|\theta_j|^n\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|x_t(\boldsymbol{\theta})|^{jn}$ and hence, if Θ is compact, we have $h(\cdot;\boldsymbol{\theta}) \in \mathbb{M}^0_{\Theta,\Theta}(n,m)$ with $m = n/J$ because $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|x_t(\boldsymbol{\theta})|^n < \infty \Rightarrow \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|h(x_t(\boldsymbol{\theta});\boldsymbol{\theta})|^{n/J} < \infty$ and $h^{(k)}(\cdot;\boldsymbol{\theta}) \in \mathbb{M}^k_{\Theta,\Theta}(n,m)$ with $n = m/(J-k) \ \forall \ (\boldsymbol{\theta},k) \in \Theta \times \mathbb{N}_0$ because $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|x_t(\boldsymbol{\theta})|^n < \infty \Rightarrow \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|h^{(k)}(x_t(\boldsymbol{\theta});\ \boldsymbol{\theta})|^{n/(J-k)} < \infty$ by the same argument.

For *(c)* we have, for some c, $\mathbb{E}|h(x_t(\boldsymbol{\theta});\boldsymbol{\theta})|^n = \mathbb{E}|\sum_{j=0}^J \theta_j x_t^{r_j}(\boldsymbol{\theta})|^n \leq c\sum_{j=0}^J \mathbb{E}|\theta_j x_t^{r_j}(\boldsymbol{\theta})|^n \leq c\sum_{j=0}^J |\theta_j|^n\mathbb{E}|x_t(\boldsymbol{\theta})|^{r_j n}$. Hence, $h(\cdot;\boldsymbol{\theta}) \in \mathbb{M}^0_{\Theta,\boldsymbol{\theta}}(n,m)$ with $m = n/\max_j r_j \ \forall \ \boldsymbol{\theta} \in \Theta$ because $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|x_t(\boldsymbol{\theta})|^n < \infty \Rightarrow \mathbb{E}|x_t(\boldsymbol{\theta})|^n < \infty \ \forall \ \boldsymbol{\theta} \in \Theta \Rightarrow \mathbb{E}|h(x_t(\boldsymbol{\theta});\boldsymbol{\theta})|^{n/\max_j r_j} \leq c\sum_{j=0}^J |\theta_j|^n\mathbb{E}|x_t(\boldsymbol{\theta})|^{r_j n/\max_j r_j} < \infty \ \forall \ \boldsymbol{\theta} \in \Theta$. Similarly, $h^{(k)}(\cdot;\boldsymbol{\theta}) \in \mathbb{M}^0_{\Theta,\boldsymbol{\theta}}(n,m)$ with $m = n/(\max_j r_j - k) \ \forall \ (\boldsymbol{\theta},k) \in \Theta \times \mathbb{N}_0 : k \leq \min_j r_j$ because we have $\mathbb{E}|h^{(k)}(x_t(\boldsymbol{\theta});\boldsymbol{\theta})|^n = \mathbb{E}|\sum_{j=0}^J \theta_j^* x_t^{r_j-k}(\boldsymbol{\theta})|^n \leq c\sum_{j=0}^J \mathbb{E}|\theta_j^* x_t^{r_j-k}(\boldsymbol{\theta})|^n \leq c\sum_{j=0}^J |\theta_j^*|^n\mathbb{E}|x_t(\boldsymbol{\theta})|^{(r_j-k)n}$ and hence it follows that $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|x_t(\boldsymbol{\theta})|^n < \infty \Rightarrow \mathbb{E}|x_t(\boldsymbol{\theta})|^n < \infty \ \forall \ \boldsymbol{\theta} \in \Theta \Rightarrow \mathbb{E}|h^{(k)}(x_t(\boldsymbol{\theta});\boldsymbol{\theta})|^{n/(\max_j r_j)} \leq c\sum_{j=0}^J |\theta_j^*|^n\mathbb{E}|x_t(\boldsymbol{\theta})|^{(r_j-k)n/(\max_j r_j - k)} < \infty$.

Furthermore, $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|h(x_t(\boldsymbol{\theta});\boldsymbol{\theta})|^n = \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|\sum_{j=0}^J \theta_j x_t^{r_j}(\boldsymbol{\theta})|^n \leq c\times$

$\sum_{j=0}^{J}\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|\theta_j x_t^{r_j}(\boldsymbol{\theta})|^n \leq c\sum_{j=0}^{J}\sup_{\boldsymbol{\theta}\in\Theta}|\theta_j|^n\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|x_t(\boldsymbol{\theta})|^{r_j n}$. Hence, if Θ is compact, we have $h \in \mathbb{M}^0_{\Theta,\Theta}(n,m)$ with $m = n/\max_j r_j$ because $\sup_{\boldsymbol{\theta}\in\Theta}|\theta_j|^n < \infty \ \forall\ j$, and hence it follows that $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|x_t(\boldsymbol{\theta})|^n < \infty \Rightarrow \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|h(x_t(\boldsymbol{\theta});\boldsymbol{\theta})|^{n/\max_j r_j} < \infty$. Similarly, we have $h^{(k)} \in \mathbb{M}^0_{\Theta,\Theta}(n,m)$ with $m = n/\max_j(r_j - k)$ because we have $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|x_t(\boldsymbol{\theta})|^n < \infty \Rightarrow \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|h(x_t(\boldsymbol{\theta});\boldsymbol{\theta})|^{n/(\max_j r_j - k)} < \infty$ by the same argument.

For *(d)* we have that $h(x_t(\boldsymbol{\theta});\boldsymbol{\theta}) \leq \bar{h}(\boldsymbol{\theta}) \ \forall\ \boldsymbol{\theta}\in\Theta \Rightarrow \mathbb{E}|h(x_t(\boldsymbol{\theta});\boldsymbol{\theta})|^n \leq \bar{h}(\boldsymbol{\theta})^n \ \forall\ (\boldsymbol{\theta},n)\in\Theta\times\mathbb{R}_0^+$, and hence, $h(\cdot;\boldsymbol{\theta}) \in \mathbb{M}^0_{\Theta,\boldsymbol{\theta}}(n,m) \forall\ (n,m,\boldsymbol{\theta}) \in \Theta\times\mathbb{R}_0^+\times\mathbb{R}_0^+$ because $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|x_t(\boldsymbol{\theta})|^n < \infty \Rightarrow \mathbb{E}|h(x_t(\boldsymbol{\theta});\boldsymbol{\theta})|^m \leq \bar{h}(\boldsymbol{\theta})^m < \infty \ \forall\ (n,m,\boldsymbol{\theta}) \in \Theta\times\mathbb{R}_0^+\times\mathbb{R}_0^+$. Furthermore, if $\sup_{\boldsymbol{\theta}\in\Theta}\bar{h}(\boldsymbol{\theta}) \leq ba\bar{r}h$ then $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|h(x_t(\boldsymbol{\theta});\boldsymbol{\theta})|^n \leq \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\bar{h}(\boldsymbol{\theta})^n \ \forall\ n\in\mathbb{R}_0^+$. Hence, $h \in \mathbb{M}^0_{\Theta,\Theta}(n,m) \ \forall\ (n,m)\in\mathbb{R}_0^+\times\mathbb{R}_0^+$ as $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|x_t(\boldsymbol{\theta})|^n < \infty \Rightarrow \sup_{\boldsymbol{\theta}\in\Theta}\mathbb{E}|h(x_t(\boldsymbol{\theta});\boldsymbol{\theta})|^m \leq \sup_{\boldsymbol{\theta}\in\Theta}\bar{h}(\boldsymbol{\theta})^m \leq \bar{\bar{h}}^m < \infty \ \forall\ (n,m)\in\mathbb{R}_0^+\times\mathbb{R}_0^+$.

For *(e)* we have, for some c, and by an exact k^{th}-order Taylor expansion around a point $x\in\text{int}(\mathcal{X})$, $\mathbb{E}|h(x_t(\boldsymbol{\theta});\boldsymbol{\theta})|^n \leq \mathbb{E}|\sum_{j=0}^{k}\theta_j x_t^j(\boldsymbol{\theta})|^n \leq c\sum_{j=0}^{J}\mathbb{E}|\theta_j x_t^j(\boldsymbol{\theta})|^n \leq c\sum_{j=0}^{J}|\theta_j|^n\mathbb{E}|x_t(\boldsymbol{\theta})|^{jn}$ where $\infty > \theta_k \geq \bar{h}_k(\boldsymbol{\theta}) \geq \sup_{x\in\mathcal{X}}|h^{(k)}(x\boldsymbol{\theta})| \ \forall\ \boldsymbol{\theta}\in\Theta$, and hence, $h(\cdot;\boldsymbol{\theta}) \in \mathbb{M}^0_{\Theta,\boldsymbol{\theta}}(n,m)$ with $m = n/k \ \forall\ \boldsymbol{\theta}\in\Theta$ because $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|x_t(\boldsymbol{\theta})|^n < \infty \Rightarrow \mathbb{E}|x_t(\boldsymbol{\theta})|^n < \infty \ \forall\ \boldsymbol{\theta}\in\Theta \Rightarrow \mathbb{E}|h(x_t(\boldsymbol{\theta});\boldsymbol{\theta})|^{n/k} \leq c\sum_{j=0}^{J}|\theta_j|^n\mathbb{E}|x_t(\boldsymbol{\theta})|^{jn/k} < \infty \ \forall\ \boldsymbol{\theta}\in\Theta$. Furthermore, $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|h(x_t(\boldsymbol{\theta});\boldsymbol{\theta})|^n = \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|\sum_{j=0}^{J}\theta_j x_t^j(\boldsymbol{\theta})|^n \leq c\sum_{j=0}^{J}\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|\theta_j\times x_t^j(\boldsymbol{\theta})|^n \leq c\sum_{j=0}^{J}\sup_{\boldsymbol{\theta}\in\Theta}|\theta_j|^n\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|x_t(\boldsymbol{\theta})|^{jn}$ and hence, if Θ is compact, we have $h(\cdot;\boldsymbol{\theta}) \in \mathbb{M}^0_{\Theta,\Theta}(n,m)$ with $m = n/k$ because $\sup_{\boldsymbol{\theta}\in\Theta}|\theta_j|^n < \infty \ \forall\ j$, and hence, $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|x_t(\boldsymbol{\theta})|^n < \infty \Rightarrow \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|h(x_t(\boldsymbol{\theta});\boldsymbol{\theta})|^{n/k} < \infty$ by a similar argument.

For *(f)* we have, for some c, $\mathbb{E}|w(x_t(\boldsymbol{\theta}),v_t(\boldsymbol{\theta});\boldsymbol{\theta})|^n = \mathbb{E}|\theta_0 + \theta_1 x_t(\boldsymbol{\theta}) + \theta_2 v_t(\boldsymbol{\theta})|^n \leq |\theta_0|^n + |\theta_1|^n\mathbb{E}|x_t(\boldsymbol{\theta})|^n + |\theta_2|^n\mathbb{E}|v_t(\boldsymbol{\theta})|^n$. Hence, $w \in \mathbb{M}^{(k_x,k_v)}_{\Theta,\boldsymbol{\theta}}(\boldsymbol{n},m) \ \forall\ (k_x,k_v)\in\mathbb{N}_0\times\mathbb{N}_0$ with $\boldsymbol{n} = (n_x,n_v)$ and $m = \min\{n_x,n_v\}$ because $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|x_t(\boldsymbol{\theta})|^{n_x} < \infty \ \wedge\ \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|v_t(\boldsymbol{\theta})|^{n_v} < \infty \Rightarrow \mathbb{E}|x_t(\boldsymbol{\theta})|^{n_x} < \infty \ \wedge\ \mathbb{E}|v_t(\boldsymbol{\theta})|^{n_v} < \infty$ implies $\mathbb{E}|w(x_t(\boldsymbol{\theta}),v_t(\boldsymbol{\theta});\boldsymbol{\theta})|^{\min\{n_x,n_v\}} \leq |\theta_0|^{\min\{n_x,n_v\}} + |\theta_1|^{\min\{n_x,n_v\}}\mathbb{E}|x_t(\boldsymbol{\theta})|^{\min\{n_x,n_v\}} + |\theta_2|^{\min\{n_x,n_v\}}\mathbb{E}|v_t(\boldsymbol{\theta})|^{\min\{n_x,n_v\}} < \infty$ and $\mathbb{E}|w^{(1,0)}(x_t(\boldsymbol{\theta}),v_t(\boldsymbol{\theta});\boldsymbol{\theta})|^{\min\{n_x,n_v\}} = |\theta_1|^n < \infty$ and similarly for v we have $\mathbb{E}|w^{(0,1)}(x_t(\boldsymbol{\theta}),v_t(\boldsymbol{\theta});\boldsymbol{\theta})|^{\min\{n_x,n_v\}} = |\theta_2|^n < \infty$ and for any derivative we have $\mathbb{E}|w^{(k_x,k_v)}(x_t(\boldsymbol{\theta}),v_t(\boldsymbol{\theta});\boldsymbol{\theta})|^{\min\{n_x,n_v\}} = 0 < \infty \ \forall\ (k_x,k_v) : k_x + k_v > 1$. Furthermore, if Θ is compact, then $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|w(x_t(\boldsymbol{\theta}),v_t(\boldsymbol{\theta});\boldsymbol{\theta})|^n = \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|\theta_0 + \theta_1 x_t(\boldsymbol{\theta}) + \theta_2 v_t(\boldsymbol{\theta})|^n \leq \sup_{\boldsymbol{\theta}\in\Theta}|\theta_0|^n + \sup_{\boldsymbol{\theta}\in\Theta}|\theta_1|^n\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|x_t(\boldsymbol{\theta})|^n + \sup_{\boldsymbol{\theta}\in\Theta}|\theta_2|^n\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|v_t(\boldsymbol{\theta})|^n$, and hence, $w \in \mathbb{M}^{(k_x,k_v)}_{\Theta,\Theta}(\boldsymbol{n},m) \ \forall\ (k_x,k_v)\in\mathbb{N}_0\times\mathbb{N}_0$ with $\boldsymbol{n} = (n_x,n_v)$ and $m = \min\{n_x,n_v\}$ because $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|x_t(\boldsymbol{\theta})|^{n_x} < \infty \ \wedge\ \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|v_t(\boldsymbol{\theta})|^{n_v} < \infty$ implies by a similar argument the bound $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|w^{(k_x,k_v)}(x_t(\boldsymbol{\theta}),v_t(\boldsymbol{\theta});\boldsymbol{\theta})|^{\min\{n_1,n_2\}}<\infty$.

For *(g)* we have $\mathbb{E}|w(x_t(\boldsymbol{\theta}),v_t(\boldsymbol{\theta});\boldsymbol{\theta})|^n<\infty$ if and only if $(\mathbb{E}|w(x_t(\boldsymbol{\theta}),v_t(\boldsymbol{\theta});\boldsymbol{\theta})|^n)^{1/n}<\infty$ and $(\mathbb{E}|w(x_t(\boldsymbol{\theta}),v_t(\boldsymbol{\theta});\boldsymbol{\theta})|^n)^{1/n}=(\mathbb{E}|\theta_0+\theta_1 x_t(\boldsymbol{\theta})v_t(\boldsymbol{\theta})|^n)^{1/n}\leq|\theta_0|+|\theta_1|(\mathbb{E}|x_t(\boldsymbol{\theta})v_t(\boldsymbol{\theta})|^n)^{1/n}\leq|\theta_0|+|\theta_1|(\mathbb{E}|x_t(\boldsymbol{\theta})|^r)^{1/r}(\mathbb{E}|v_t(\boldsymbol{\theta})|^s)^{1/s}$ with $1/r+1/s=1/n$ by the generalized Holder's inequality, and hence, $w\in\mathbb{M}^{(k_x,k_v)}_{\Theta,\boldsymbol{\theta}}(\boldsymbol{n},m)\ \forall\,(k_x,k_v)\in\mathbb{N}_0\times\mathbb{N}_0$ with $\boldsymbol{n}=(n_x,n_v)$ if $1/m=1/n_x+1/n_v\Leftrightarrow m=n_xn_v/(n_x+n_v)$ because then $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|x_t(\boldsymbol{\theta})|^{n_x}<\infty\ \wedge\ \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|v_t(\boldsymbol{\theta})|^{n_v}<\infty$ and this implies $\mathbb{E}|x_t(\boldsymbol{\theta})|^{n_x}<\infty\ \wedge\ \mathbb{E}|v_t(\boldsymbol{\theta})|^{n_v}<\infty\Rightarrow\mathbb{E}|w(x_t(\boldsymbol{\theta}),v_t(\boldsymbol{\theta});\boldsymbol{\theta})|^{\frac{n_xn_v}{n_x+n_v}}<\infty$. Furthermore, if Θ is compact, then $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|w(x_t(\boldsymbol{\theta}),v_t(\boldsymbol{\theta});\boldsymbol{\theta})|^n<\infty$ iff $(\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|w(x_t(\boldsymbol{\theta}),v_t(\boldsymbol{\theta});\boldsymbol{\theta})|^n)^{1/n}<\infty$ and since we have $(\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|w(x_t(\boldsymbol{\theta}),v_t(\boldsymbol{\theta});\boldsymbol{\theta})|^n)^{1/n}=(\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|\theta_0+\theta_1x_t(\boldsymbol{\theta})v_t(\boldsymbol{\theta})|^n)^{1/n}\leq\sup_{\boldsymbol{\theta}\in\Theta}|\theta_0|+\sup_{\boldsymbol{\theta}\in\Theta}|\theta_1|(\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|x_t(\boldsymbol{\theta})v_t(\boldsymbol{\theta})|^n)^{1/n}\leq\sup_{\boldsymbol{\theta}\in\Theta}|\theta_0|+\sup_{\boldsymbol{\theta}\in\Theta}|\theta_1|(\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|x_t(\boldsymbol{\theta})|^r)^{1/r}(\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|v_t(\boldsymbol{\theta})|^s)^{1/s}$ with r and s satisfying $1/r+1/s=1/n$ by the generalized Holder's inequality, and hence, $w\in\mathbb{M}^{(k_x,k_v)}_{\Theta,\Theta}(\boldsymbol{n},m)\ \forall\,(k_x,k_v)\in\mathbb{N}_0\times\mathbb{N}_0$ with $\boldsymbol{n}=(n_x,n_v)$ if $m=n_xn_v/(n_x+n_v)$ by a similar argument. □

F Additional GAS Illustrations

F.1 Example 2: Dynamic one-factor model

Let y_{it} denote the ith time series in a panel of dimension i, for $i=1,\ldots,d_y$. Each time series is modeled by

$$y_{it}=a_i+b_if_t+c_iu_{it},\qquad i=1,\ldots,d_y,\tag{F.1}$$

where $a_i=a_i(\lambda)$, $b_i=b_i(\lambda)$ and $c_i=c_i(\lambda)$ are fixed and known functions of λ only and p_u is the standard normal density. Equation (F.1) can be viewed as an observation-driven dynamic one-factor model. The GAS transition equation is given by

$$f_{t+1}=\omega+\alpha(y_t^*-f_t)+\beta f_t,\qquad y_t^*=\frac{\sum_{i=1}^{d_y}b_i(y_{it}-a_i)\,/\,c_i^2}{\sum_{i=1}^{d_y}b_i^2\,/\,c_i^2},$$

where the scaling $S(f_t;\lambda)$ is equal to the inverse conditional variance of the score. Applications of dynamic one-factor models can be found in the literature on modelling interest rates y_{it} for different maturities, see Vasicek (1977), or modelling mortality rates for different age cohorts i, see Lee and Carter (1992).

F.2 Example 3: Conditional duration models

If y_t is strictly positive, we can set $g(f_t, u_t) = f_t u_t$ and choose p_u as a positively valued random variable with mean 1. For example, let u_t have a Gamma distribution with mean 1 and variance λ^{-1}. Scaling the conditional score by its conditional variance, we obtain

$$f_{t+1} = \omega + \alpha(y_t - f_t) + \beta f_t, \tag{F.2}$$

which reduces to the MEM(1,1) model of Engle (2002) with the autoregressive conditional duration (ACD) model of Engle and Russell (1998) as a special case ($\lambda = 1$). We notice that the GAS model for $g(f_t, u_t) = f_t u_t$ with p_u a Gamma density is the same as the GAS model for $g(f_t, u_t) = \log(f_t) + u_t$ with $\exp(u_t)$ a Gamma distributed random variable. A transformation of variables for y_t that is independent of f_t thus leaves the GAS model unaffected. If p_u is a fat-tailed distribution such as a Gamma mixture of exponentials, $p_u(u_t; \lambda) = (1+\lambda^{-1}u_t)^{-(1+\lambda)}$ for $\lambda > 0$, we obtain under an appropriate choice of the scaling function the recursion

$$f_{t+1} = \omega + \alpha \left(\frac{(1 + \lambda^{-1})y_t}{1 + \lambda^{-1} y_t / f_t} - f_t \right) + \beta f_t, \tag{F.3}$$

see Koopman et al. (2012) and Harvey (2013). As in Example 1, large values of y_t in (F.3) have a reduced impact on future values f_{t+1} due to the recognition that p_u is fat-tailed for $\lambda^{-1} > 0$.

F.3 Example 4: Regression with time-varying constant

To illustrate the construction of a time-varying constant for a regression model in our GAS setting of Section 2, we let p_u be the normal density with standard deviation $\lambda > 0$ and we assume $g(f_t, u_t) = f_t + X_t\delta + u_t$ where X_t is a row vector of exogenous or conditionally determined variables and δ is a column vector of fixed coefficients. We obtain the following nonlinear conditional time-varying regression model

$$y_t = f_t + X_t\delta + u_t, \qquad u_t \sim \mathrm{N}(0, \lambda^2). \tag{F.4}$$

The GAS updating equation for the time-varying constant f_t is given by

$$f_{t+1} = \omega + \alpha[(y_t - X_t\delta) - f_t] + \beta f_t,$$

for which we have set the scaling $S(f_t; \lambda)$ equal to the information matrix with respect to f_t. The unknown coefficient vector δ is linear in y_t and can typically be concentrated out from the likelihood function. See also Harvey and Luati (2014) for fat-tailed extensions of this model.

G Additional Applications of the Theory to GAS Models

G.1 Further Details on Time-Varying Mean for the Skewed Normal

In this example, $\tilde{\rho}_t^k(\boldsymbol{\theta})$ is calculated as

$$\begin{aligned}\tilde{\rho}_t^k(\boldsymbol{\theta}) &= \sup_{f^* \in \mathcal{F}^*} \left| \beta + \alpha \left(-1 + \lambda^2 \, \frac{\partial}{\partial z} \frac{z \, p_{\mathrm{N}}(z)^2}{P_{\mathrm{N}}(z)} \bigg|_{z=\lambda(y_t - f^*)} \right) \right|^k \\ &\approx \max\left(|\beta - \alpha\,(1 - 0.436\lambda^2)| \,, \; |\beta - \alpha\,(1 + 0.289\lambda^2)| \right)^k. \qquad \text{(G.1)}\end{aligned}$$

G.2 Further Details on Student's t Time Varying Conditional Volatility Models

$$S(f_t;\lambda) = \mathcal{I}^{-1}(f_t;\lambda) = 2(1+3\lambda^{-1})f_t^2, \qquad \text{(G.2)}$$

$$s(f_t, y_t;\lambda) = \left(1+3\lambda^{-1}\right)\left(\frac{(1+\lambda^{-1})y_t^2}{1+y_t^2/(\lambda f_t)} - f_t\right), \qquad \text{(G.3)}$$

$$\dot{s}_{y,t}(f_t;\lambda) = \left(1+3\lambda^{-1}\right)\left(\frac{(1+\lambda)(y_t^2/(\lambda f_t))^2}{(1+y_t^2/(\lambda f_t))^2} - 1\right), \qquad \text{(G.4)}$$

$$s_u(f_t, u_t;\lambda) = \dot{s}_{u,t}(f_t;\lambda)\cdot f_t, \qquad \text{(G.5)}$$

$$\dot{s}_{u,t}(f_t;\lambda) = \left(1+3\lambda^{-1}\right)\left(\frac{(1+\lambda^{-1})u_t^2}{1+\lambda^{-1}u_t^2} - 1\right). \qquad \text{(G.6)}$$

To ensure that f_t is always positive, we require $\beta > (1+3\lambda^{-1})\alpha > 0$ and $\omega \geq \underline{\omega} > 0$. We also define $\alpha^* = \alpha^*(\alpha,\lambda) = (1+3\lambda^{-1})(1+\lambda^{-1})\alpha$ and $\beta^* = \beta^*(\alpha,\beta,\lambda) := \beta - (1+3\lambda^{-1})\alpha > 0$ and assume $\beta^* < 1$, such that f_t converges exponentially fast to $\underline{f} = \underline{f}(\boldsymbol{\theta}) := \omega/(1-\beta^*)$ if we set $y_t, y_{t+1}, \ldots$ all equal to zero.

An estimate of the region for global consistency and asymptotic normality under mis-specification is obtained as follows by estimating the expectation of

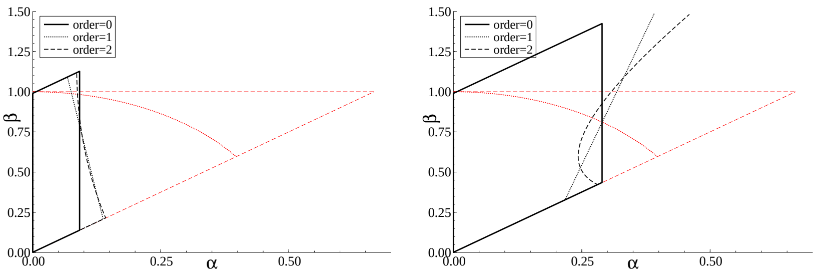

Figure 1: Local and global consistency and asymptotic normality regions for the Student's t GAS model. Local consistency under correct specification is obtained in the triangle, and local asymptotic normality below the dashed curve in the triangle. Global consistency and asymptotic normality are established in the regions bounded by fat-lines. The boundary of the regions is approximated by a polynomial in β of specific order. The left-panel is for 1,000 simulated observations of a Student's t GAS model with $\beta = 0.80$ and $\alpha = 0.05$. The right-hand panel is for $\beta = \alpha = 0$.

$\sup_{\boldsymbol{\theta}\in\Theta} \log \tilde{\rho}_t^k(\boldsymbol{\theta})$ by the sample average,

$$\frac{1}{T}\sum_{t=1}^{T} \sup_{\boldsymbol{\theta}\in\Theta} \log \left(\beta^* + \alpha\,(1+3\lambda^{-1}) \frac{(1+\lambda)(y_t^2/(\lambda \underline{f}))^2}{(1+y_t^2/(\lambda \underline{f}))^2} \right) < 0. \tag{G.7}$$

From the uniform boundedness of the score and the assumption that $\{y_t\}$ is SE, we establish that the left-hand side of (G.7) is a consistent estimator for the expectation. Moreover, using Remark 4 we directly obtain the existence of appropriate moments once we establish the SE condition. The main advantage of using the sample average as an estimator of the expectation is that it automatically puts more weight on the relevant area of the sample space for a given process $\{y_t\}$.

There are different parameter spaces Θ that satisfy (G.7). For a parameter space with polynomial boundary $\bar{\beta}(\alpha) = \sum_{i=1}^{q} c_i\,\alpha^i$ for $q = 0, 1, 2$, the results are presented for $\omega = 1$ and $\lambda = 6$ in Figure 1 by the bold curves. Comparing the left-hand and right-hand panels, we note that the regions become larger for less persistent processes.

G.3 Logistic Tracking with Fat Tailed Innovations

Consider a robust, nonlinear model setting where outliers in $\{y_t\}_{t\in\mathbb{Z}}$ are generated by fat tailed i.i.d. Student's t innovations rather than by abrupt changes in the conditional mean, i.e., $y_t = h(f_t) + u_t$. We assume $u_t \sim t(0, \lambda_1, \lambda_2)$, where 0, λ_1, and λ_2 are the location, scale, and degrees of freedom parameter of the Student's t distribution, respectively. We consider a GAS model where the conditional mean $h(f_t)$ is given by a logistic mapping $h(f_t) = (1+e^{-f_t})^{-1}$, such that the mean is constrained between 0 and 1 by construction. We use constant scaling $S(f_t; \lambda) = \lambda_1(1+\lambda_2^{-1})^{-1}$, which slightly simplifies the expression for the scaled score later on. The GAS update now tracks the conditional expectation indirectly through $\{f_t\}$.

A simulated example is presented in Figure 2 for 250 observations from a Student's t distribution with 3 degrees of freedom. The figure presents the data, which obviously include a number of tail observations. The figure also holds the true f_t sequence, and two estimated versions. All of these clearly lie in the [0,1] interval by construction due to the logistic transformation of the mean. One of the fitted f_t sequences corresponds to a mis-specified Gaussian GAS model for the conditional mean. The other corresponds to a correctly specified Student's t GAS model. As with the volatility example in the introduction, the sensitivity of the Gaussian model for tail observations is obvious; see the circled areas in the figure. The Student's t model does not suffer from this problem and provides a close fit to the true f_t sequence at all times.

We have $\tilde{g}(f_t, y_t) = y_t - (1+e^{-f_t})^{-1}$ and the first two derivatives of $h(f_t)$ with respect to f_t are equal to $h'(f_t) = e^{f_t}(1+e^{f_t})^{-2}$ and $h''(f_t) = e^{f_t}(1-e^{f_t})(1+e^{f_t})^{-3}$, respectively. Define $\lambda_3 = \lambda_1\lambda_2$, then we also have $\log \tilde{g}'(f_t, y_t) = 0$ and $\tilde{p}(u_t; \lambda_1, \lambda_2) = \log(\Gamma((\lambda_2+1)/2)/\Gamma(\lambda_2/2)) - 0.5\log(\pi\lambda_3) - 0.5(\lambda_2+1)\log(1+u_t^2/\lambda_3)$ for $\lambda_1, \lambda_2, \lambda_3 > 0$, such that

$$s(f_t, y_t; \lambda) = \frac{e^{f_t}}{(1+e^{f_t})^2} \cdot \frac{y_t - (1+e^{-f_t})^{-1}}{1+\lambda_3^{-1}(y_t - (1+e^{-f_t})^{-1})^2}. \tag{G.8}$$

Even though the score is now a complex nonlinear function of the dynamic parameter f_t, we can still use the theory developed in the previous sections. The key lies in the observation that the scaled score s and a number of its derivatives are uniformly bounded in both y_t and f_t on a compact parameter space Θ.

To establish existence and consistency of the MLE, we need to verify Assumptions 1–4. Assumptions 1 and 2 are trivially satisfied. The first part of Assumption 3 is satisfied for arbitrarily large $n_f > 0$. This follows by the fact that the first factor in (G.8) is uniformly bounded by 0.25, and the second factor is uniformly bounded by $0.5\lambda_3^{1/2}$. As a result, the score s is uniformly bounded

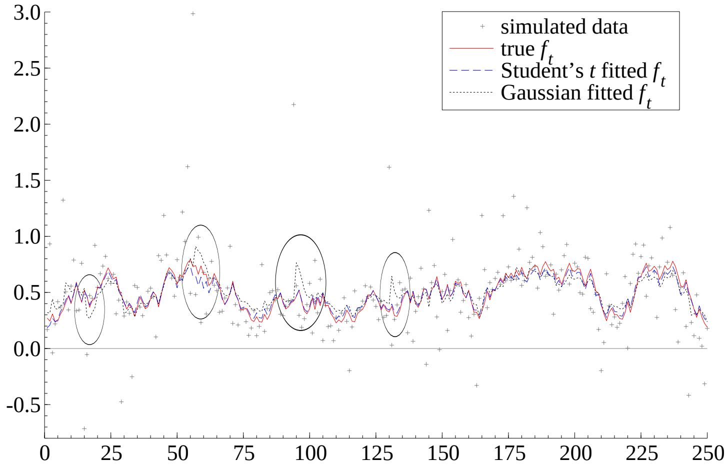

Figure 2: Simulated data $y_t = (1 + e^{-f_t})^{-1} + u_t$, with $u_t \sim t(0, 0.2^2, 3)$, and fitted f_t sequences based on a mis-specified Gaussian and a correctly specified Student's t GAS model for the dynamic (logistically constrained) conditional mean specification. The data generating process has $\omega = 0$, $\alpha = 10$, and $\beta = 0.95$.

by $0.125\lambda_3^{1/2}$, which is finite on a compact set $\Lambda \subset \mathbb{R}^+ \times \mathbb{R}^+$. Similarly, the derivative of s with respect to f is uniformly bounded, such that condition (i) of Assumption 3 holds. Also Assumption 4 holds with $n_{\tilde{g}} = n_y$ and any (large) $n_{\tilde{p}}$ and $n_{\log \tilde{g}'}$, such that $n_\ell \geq 1$ is easily satisfied as long as there is some small positive moment n_y of y_t. This establishes the existence and strong consistency of the MLE.

Under an axiom of correct specification, we can again appeal to Theorem 3 and Corollary 1 to obtain the consistency of the MLE under primitive conditions that ensure uniqueness of $\boldsymbol{\theta}_0$ and appropriate properties for $\{y_t\}_{t\in\mathbb{Z}}$. We notice that Assumption 5 holds for $0 < n_u = n_y < \lambda_2$, and n_f arbitrarily large. Also Assumption 6 holds with $\zeta(u_t;\lambda) = u_t/(1+\lambda_3^{-1}u_t^2)$ and $\eta(f_t;\lambda) = e^{f_t}(1+e^{f_t})^{-2}$, such that $\mathbb{E}|\zeta(u_t;\lambda)|^2 < \lambda_3^{1/2}/2 < \infty$ by the uniform boundedness of $|\zeta(u_t;\lambda)|$.

Using the above derivations, it follows immediately that the simple condition stated in Remark 5 requiring $m \geq 6$ moments for asymptotic normality of the MLE is too restrictive. Therefore, we verify the more elaborate moment conditions of Assumption 7 directly. First, it is straightforward to verify that $n_f^{(1)}$ and $n_f^{(2)}$ can both be taken arbitrarily large due to the boundedness of s and its derivatives. It, however, may not be very useful as large values of n_f

can shrink the region Θ^*. To derive the minimum values of $n_f^{(1)}$ and $n_f^{(2)}$ for which Assumption 7 holds, we first emphasize that we can set m_B arbitrarily large because we have the uniform boundedness of the derivatives of $\tilde{p}(\tilde{g};\lambda)$ with respect of $\tilde{g}$. Moreover, as $\tilde{p}^{(0,1)}$ is at most logarithmic in $(y_t-(1+e^{-f_t})^{-1})^2$ and $\tilde{p}^{(0,2)}$ is uniformly bounded, we can also set $n_{\tilde{p}}^{(0,1)}$ and $n_{\tilde{p}}^{(0,2)}$ arbitrarily large as long as a small moment $n_y > 0$ of y_t exists. Assumption 7 is therefore satisfied as long as $n_f = n_f^{(1)} = n_f^{(2)} = 2$.

The corresponding region where asymptotic normality can be ensured is characterized by

$$\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\tilde{\rho}_t^2(\boldsymbol{\theta}) = \mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\sup_{f^*\in\mathcal{F}^*}\left|\beta+\alpha\,\dot{s}_{y,t}(f^*;\lambda)\right|^2 < 1. \tag{G.9}$$

This expression is highly complex and there are two different ways to verify for which parameter combinations it is satisfied: an analytic way by obtaining a uniform bound, and a numerical way by obtaining an estimate of the expectation in (G.9). We illustrated the latter approach in Section G.2. Here, we illustrate the approach based on a uniform bound and by using Remark 4 from Section 3.

With $\lambda_3 = \lambda_1\lambda_2$, we have

$$\begin{aligned}
&\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\sup_{f^*\in\mathcal{F}^*}\left|\beta+\alpha\,\dot{s}_{y,t}(f^*;\lambda)\right|^2 \leq \\
&\sup_{\boldsymbol{\theta}\in\Theta}\Big(|\beta|+|\alpha|\sup_{u\in\mathbb{R}}\sup_{f^*\in\mathcal{F}^*}\Big|\frac{-e^{2f^*}}{(1+e^{f^*})^4}\frac{1-\lambda_3^{-1}u^2}{(1+\lambda_3^{-1}u^2)^2}+ \\
&\qquad\qquad\frac{e^{f^*}(1-e^{f^*})}{(1+e^{f^*})^3}\frac{u}{1+\lambda_3^{-1}u^2}\Big|\Big)^2 \leq \\
&\sup_{\boldsymbol{\theta}\in\Theta}\Big(|\beta|+|\alpha|\sup_{u\in\mathbb{R}}\Big(\frac{1}{4}\Big|\frac{1-\lambda_3^{-1}u^2}{(1+\lambda_3^{-1}u^2)^2}\Big|+\big(\frac{1}{2}+\frac{1}{6}\sqrt{3}\big)\Big|\frac{u}{1+\lambda_3^{-1}u^2}\Big|\Big)\Big)^2 \leq \\
&\sup_{\boldsymbol{\theta}\in\Theta}\Big(|\beta|+|\alpha|\Big(\frac{1}{4}+\big(\frac{1}{2}+\frac{1}{6}\sqrt{3}\big)\frac{\sqrt{\lambda_3}}{2}\Big)\Big)^2 = \\
&\sup_{\boldsymbol{\theta}\in\Theta}\Big(|\beta|+|\alpha|\Big(\frac{1}{4}+\big(\frac{1}{4}+\frac{1}{12}\sqrt{3}\big)\sqrt{\lambda_3}\Big)\Big)^2 < 1.
\end{aligned}$$

Under an axiom of correct specification, the results remain largely unaltered. The main difference lies in the fact that Θ has to be a subset of the intersection of Θ^* with Θ_*^* as defined in Assumption 5. Using the same approach with a uniform bound as for the scaled score expression in y_t, we obtain

$$\begin{aligned}
&\mathbb{E} \sup_{\boldsymbol{\theta}\in\Theta} \sup_{f^*\in\mathcal{F}^*} \Big|\beta + \alpha\, \dot{s}_{u,t}(f^*;\lambda)\Big|^2 \\
&\quad \leq \mathbb{E} \sup_{\boldsymbol{\theta}\in\Theta} \sup_{f^*\in\mathcal{F}^*} \Big|\beta - \alpha\, \frac{e^{f^*}(1-e^{f^*})}{(1+e^{f^*})^3} \frac{u_t}{1+\lambda_3^{-1}u_t^2}\Big|^2 \\
&\quad \leq \sup_{\boldsymbol{\theta}\in\Theta} \Big(|\beta| + |\alpha| \sup_{u\in\mathbb{R}} \sup_{f^*\in\mathcal{F}^*} \frac{e^{f^*}(1-e^{f^*})}{(1+e^{f^*})^3} \frac{u}{1+\lambda_3^{-1}u^2}\Big)^2 \\
&\quad \leq \sup_{\boldsymbol{\theta}\in\Theta} \Big(|\beta| + |\alpha| \sup_{u\in\mathbb{R}} \frac{3\sqrt{3}-5}{54-30\sqrt{3}} \frac{u}{1+\lambda_3^{-1}u^2}\Big)^2 \\
&\quad \leq \sup_{\boldsymbol{\theta}\in\Theta} \Big(|\beta| + |\alpha| \frac{3\sqrt{3}-5}{108-60\sqrt{3}} \sqrt{\lambda_3}\Big)^2 \\
&\quad = \sup_{\boldsymbol{\theta}\in\Theta} \Big(|\beta| + \frac{|\alpha|\sqrt{3}}{36} \sqrt{\lambda_3}\Big)^2 < 1.
\end{aligned}$$

The sufficient region Θ^* as derived above for the mis-specified model is obviously smaller than that for the correctly specified model, Θ^*_*. The sufficient regions have a diamond shape and are decreasing in λ_3. As mentioned earlier, however, these analytic bounds derived from the uniform bound on the score function may in many cases be rather strict. An alternative way to check the bounds would be by means of an empirical estimate of the expectation in (G.9) as in Section G.2.

Acknowledgements

We thank Peter Boswijk, Christian Francq, Andrew Harvey, and Anders Rahbek, as well as the participants of the "2013 Workshop on Dynamic Models driven by the Score of Predictive Likelihoods", Amsterdam; the "7th International Conference on Computational and Financial Econometrics (CFE 2013)", London; and the "2014 Workshop on Dynamic Models driven by the Score of Predictive Likelihoods", Tenerife, for helpful comments and discussions.

References

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

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. (2013). *Dynamic Models for Volatility and Heavy Tails.* Cambridge University Press.

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

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*.

Krengel, U. (1985). *Ergodic theorems.* Berlin: De Gruyter studies in Mathematics.

Lee, R. D. and L. Carter (1992). Modeling and forecasting the time series of u.s. mortality. *Journal of the American Statistical Association 87*(Sept), 659–671.

Loève, M. (1977). *Probability theory.* Springer-Verlag, New York.

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

Vasicek, O. (1977). An equilibrium characterisation of the term structure. *Journal of Financial Economics 5*(2), 177–188.