TI 2014-074/III
Tinbergen Institute Discussion Paper

Maximum Likelihood Estimation for correctly Specified Generalized Autoregressive Score Models: Feedback Effects, Contraction Conditions and Asymptotic Properties

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

[a] *Faculty of Economics and Business Administration, VU University Amsterdam, and Tinbergen Institute, the Netherlands;*
[b] *CREATES, Aarhus University, Denmark.*

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

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

Tinbergen Institute has two locations:

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

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

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

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

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

Maximum Likelihood Estimation for Correctly Specified Generalized Autoregressive Score Models:

Feedback Effects, Contraction Conditions and Asymptotic Properties*

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

(a) *VU University Amsterdam and Tinbergen Institute*
(b) *CREATES, Aarhus University*

June 16, 2014

Abstract

The strong consistency and asymptotic normality of the maximum likelihood estimator in observation-driven models usually requires the study of the model both as a filter for the time-varying parameter and as a data generating process (DGP) for observed data. The probabilistic properties of the filter can be substantially different from those of the DGP. This difference is particularly relevant for recently developed time varying parameter models. We establish new conditions under which the dynamic properties of the true time varying parameter as well as of its filtered counterpart are both well-behaved and We only require the verification of one rather than two sets of conditions. In particular, we formulate conditions under which the (local) invertibility of the model follows directly from the stable behavior of the true time varying parameter. We use these results to prove the local strong consistency and asymptotic normality of the maximum likelihood estimator. To illustrate the results, we apply the theory to a number of empirically relevant models.

Keywords: Observation-driven models, stochastic recurrence equations, contraction conditions, invertibility, stationarity, ergodicity, generalized autoregressive score models.

1 Introduction

For a general class of observation-driven time varying parameter models, we present two contributions with respect to the local asymptotic properties of the maximum likelihood estimator (MLE). First, we show that, under appropriate conditions, local model invertibility can be obtained as a by-product of establishing the stationarity and ergodicity of the data generating process (DGP) through a contraction condition. We show that, depending on the nature of the model, these conditions may be very mild or quite restrictive. Second, we explore the theory developed in Blasques et al.

*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. Blasques and Lucas thank the Dutch National Science Foundation (NWO; grant VICI453-09-005) for financial support. Koopman acknowledges support from CREATES, Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation. Email correspondence: f.blasques@vu.nl, s.j.koopman@vu.nl, a.lucas@vu.nl.

(2014b) and prove their unsubstantiated claims concerning the local strong consistency and asymptotic normality of the MLE for the parameters of a correctly-specified generalized autoregressive score model. We analyze the underlying conditions and the extent to which they may be restrictive.

We present our results for the class of generalized autoregressive score (GAS) models which are developed by Creal et al. (2011, 2013) and Harvey (2013). In the GAS framework, model parameters are made time varying through the score of the predictive observation density function. The GAS model encompasses well-known observation driven time varying parameter models including the generalized autoregressive conditional heteroskedasticity (GARCH) model of Engle (1982) and Bollerslev (1986) and the autoregressive conditional duration (ACD) model of Engle and Russell (1998). It further gives rise to new interesting models such as the observation-driven mixed measurement dynamic factor models of Creal et al. (2014), the dynamic models for location, volatility, and multivariate dependence for fat-tailed densities of Creal et al. (2011), Lucas et al. (2014), Harvey and Luati (2014), Andres (2014), and many more. A complete compilation of GAS-model related work can be found online at `http://gasmodel.com`.

For an observation y_t with density $p_y(y_t|f_t;\lambda)$, time-varying parameter f_t and fixed static parameter vector λ, the GAS model determines the next value f_{t+1} based on the current values for f_t and y_t. The GAS model is an observation-driven model as defined in Cox (1981). The defining feature of GAS is its use of the score function of the conditional or predictive observation density in the parameter updating. The basic GAS model is specified as

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

where $s(f_t, y_t;\lambda)$ is the scaled score of the predictive density p_y,

$$s(f_t, y_t;\lambda) = S(f_t;\lambda)\cdot \partial \log p_y(y_t|f_t;\lambda)/\partial f_t,$$

with some positive scaling function $S(f_t;\lambda)$ and fixed static parameters ω, α and β. The use of the score function as a driving mechanism for the time varying parameter f_t makes the GAS framework generically applicable whenever one is willing to assume a specification for the conditional density p_y. For example, if p_y is a normal distribution with mean zero and variance f_t, and if we set $S(f_t;\lambda) = 2f_t^2$, i.e. the inverse conditional Fisher information of f_t as in Creal et al. (2013), we obtain

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

with $\alpha^* = \alpha$ and $\beta^* = \beta - \alpha$. We easily recognize (2) as the familiar GARCH model of Engle (1982) and Bollerslev (1986); see for example Straumann (2005), Straumann and Mikosch (2006) and Francq and Zakoïan (2010) for overviews of asymptotic results for the GARCH model. We refer to Creal et al. (2013) and Harvey (2013) for more details on and examples of GAS models.

The likelihood function for the GAS model is known in closed form by means of a prediction error decomposition. This facilitates parameter estimation via the method of maximum likelihood. The theoretical properties of the MLE for GAS models have been explored recently by Harvey (2013) and Blasques et al. (2014b). The optimality properties of using the score function in (1) are investigated in Blasques et al. (2014a). Theoretical work for the GAS model is complicated given its dual intepretation as a data generating process for the observations y_t and as an estimation method for the latent time varying parameter process. In the former case we refer to (1) as a GAS process and in the latter case as a GAS filter.

We explore the theory further by providing insight into the intricate relationship between the stability conditions for the GAS *process* versus the GAS *filter*. The conditions for the GAS process are needed to investigate the properties of y_t. The conditions for the GAS filter are needed to establish the properties of the time varying parameter estimates evaluated at a static parameter value unequal to the true static parameter. The filtered estimates also determine the value of the likelihood function. The filtering problem is closely linked to the issue of model invertibility; see the elaborate discussion in Wintenberger (2013). To establish consistency and asymptotic normality of the MLE, we require two separate sets of conditions corresponding to two sets of nonlinear dynamic systems: the process and the filter.

In our current study we formulate a new set of conditions under which the above two sets of separate conditions become (locally) equivalent. Model invertibility can then be obtained as a by-product of establishing stationarity and ergodicity of the GAS process evaluated at the true parameter value. These results further allow us to formulate the conditions under which we obtain local strong consistency and asymptotic normality results which are provided in Blasques et al. (2014b). We find that the conditions needed for equivalence of the stability of the GAS filter and of the GAS process can be rather restrictive, but still applicable to a number of relevant empirical models. We further notice that the differences between the contraction conditions for the filter and the process are not always sharply distinguished and sufficiently highlighted in the current literature; see for example the discussion in Harvey and Luati (2014). This is particularly relevant within the class of GAS models because the nature of the nonlinearity for GAS filtering recursions can be very different from the nonlinearity of the GAS data generating process.

The results in our study rely heavily on the contraction approach for stochastic recurrence equations as developed by Bougerol (1993) and Straumann and Mikosch (2006). The remainder of this paper is organized as follows. In Section 2, we introduce the main framework for our analysis and highlight the main differences between obtaining local versus global results for observation-driven time series models. In Section 3, we formulate the asymptotic properties of the GAS process and of the MLE of the static parameters in the model. In Section 4, we provide three key examples illustrating the vastly different implications of considering either the process or filter dynamics. Section 5 concludes, and the Appendix gathers the proofs. A supplemental technical appendix is available with some of the more technical derivations.

2 Stochastic Properties of Observation-Driven Models

2.1 The Observation-Driven Model as a Data Generating Process

Under an axiom of correct specification, there exists a $\boldsymbol{\theta}_0 \in \Theta$ such that observed data $\{y_t\}_{t=1}^T$ is a subset of the realized path of a stochastic sequence $\{y_t\}_{t\in\mathbb{Z}}$ generated by

$$y_t = g(f_t(\boldsymbol{\theta}_0), u_t), \quad u_t \sim p_u(\,\cdot\,; \boldsymbol{\theta}_0), \tag{3}$$

$$f_t(\boldsymbol{\theta}_0) = \phi_y(f_t(\boldsymbol{\theta}_0), y_t; \boldsymbol{\theta}_0) \;\forall\; t \in \mathbb{Z}, \tag{4}$$

where $\{f_t(\boldsymbol{\theta}_0)\}_{t\in\mathbb{Z}}$ is the true sequence of time varying parameters with elements taking values in a convex set $\mathcal{F} \subset \mathbb{R}$; $\{u_t\}_{t\in\mathbb{Z}}$ is a sequence of independently and identically distributed (i.i.d.) innovations; f_t characterizes the conditional density of $\{y_t\}_{t\in\mathbb{Z}}$ via the function g in the observation equation (3); and ϕ_y is the function that defines the dynamic update equation for $f_t(\boldsymbol{\theta}_0)$, Whenever convenient, we drop the argument of $f_t(\boldsymbol{\theta}_0)$ and write f_t instead. As an example, consider the GAS model in (1). We have $\phi_y(f_t, y_t; \boldsymbol{\theta}_0) = \omega_0 + \alpha_0 s(f_t, y_t; \lambda_0) + \beta_0 f_t$. In particular for the GARCH model in equation (2), we obtain $g(f_t, u_t) = (f_t)^{1/2} u_t$, $u_t \sim N(0, \sigma_0^2)$, and $f_{t+1} = \phi_y(f_t, y_t; \boldsymbol{\theta}_0) = \omega_0 + \alpha_0 y_t^2 + \beta_0 f_t$.

In order to analyze the stochastic properties of the true time-varying parameter $\{f_t(\boldsymbol{\theta}_0)\}_{t\in\mathbb{Z}}$ it is important to work with both equations (3) and (4) so as to recognize that changes in $f_t(\boldsymbol{\theta}_0)$ affect the data y_t through (3), which in turn feed back into $f_{t+1}(\boldsymbol{\theta}_0)$ through the update equation (4). This feedback is conveniently analyzed by re-writing the recursion for $f_t(\boldsymbol{\theta}_0)$ in (4) in terms of the innovations u_t only,

$$f_{t+1}(\boldsymbol{\theta}_0) = \phi_u\big(f_t(\boldsymbol{\theta}_0), u_t; \boldsymbol{\theta}_0\big) := \phi_y\big(f_t(\boldsymbol{\theta}_0), g(f_t(\boldsymbol{\theta}_0), u_t); \boldsymbol{\theta}_0\big) \quad \forall\; t \in \mathbb{Z}. \tag{5}$$

For example, for the GARCH model (2), the recursion in (5) takes the form

$$f_{t+1} = \phi_u(f_t, u_t; \boldsymbol{\theta}_0) = \omega_0 + (\beta_0 + \alpha_0 u_t^2) f_t.$$

The recursive form in (5) thus plays a central role in analyzing the properties of the true sequence $\{f_t(\boldsymbol{\theta}_0)\}_{t\in\mathbb{Z}}$ and those of the data $\{y_t\}_{t\in\mathbb{Z}}$ generated by the observation-driven model. For example, when ϕ_u is differentiable, then the following log moment and contraction conditions render the process $\{f_{t+1}(\boldsymbol{\theta}_0)\}_{t\in\mathbb{Z}}$ strictly stationary and ergodic: (SE)

$$\mathbb{E}\log^+|\phi_u(f, u_t; \boldsymbol{\theta}_0)| < \infty, \qquad \mathbb{E}\log\sup_f |\phi_u'(f, u_t; \boldsymbol{\theta}_0)| < 0,$$

where $\phi_u'(f, u_t; \boldsymbol{\theta}_0) := \partial\phi_u(f, u_t; \boldsymbol{\theta}_0)/\partial f$; see e.g. Bougerol (1993) and Straumann and Mikosch (2006).[1] If g is continuous, then it follows also by Krengel (1985) that $\{y_t\}_{t\in\mathbb{Z}}$ is SE.

2.2 The Observation-Driven Model as a Filter

When we are interested in analyzing the properties of the model as a filter for the true unknown time-varying parameter $\{f_t(\boldsymbol{\theta}_0)\}_{t\in\mathbb{Z}}$, then the analysis focuses essentially

[1] Alternative conditions are found in the geometric ergodicity literature; see e.g. Cline and Pu (1999) and Meyn and Tweedie (2009).

on the ϕ_y recursion in (4). Let $\{\tilde{f}_t(\boldsymbol{\theta}, \bar{f}_1)\}_{t\in\mathbb{N}}$ denote the filtered dynamic parameter evaluated at $\boldsymbol{\theta}$ and initialized at $\bar{f}_1 \in \mathcal{F} \subseteq \mathbb{R}$. Then the postulated data generating process in (4) suggests the following filtering equation for $\{\tilde{f}_t(\boldsymbol{\theta}, \bar{f}_1)\}_{t\in\mathbb{N}}$,

$$\tilde{f}_{t+1}(\boldsymbol{\theta}, \bar{f}_1) = \phi_y\big(\tilde{f}_t(\boldsymbol{\theta}, \bar{f}_1), y_t; \boldsymbol{\theta}\big) \quad \forall\, t \in \mathbb{N}. \tag{4$'$}$$

Recall that a random variable $\tilde{f}_t(\boldsymbol{\theta}, \bar{f}_1)$ converges exponentially fast almost surely (e.a.s.) to a random variable $\tilde{f}_t(\boldsymbol{\theta})$ if there exists a $c > 1$ such that $c^t\,|\tilde{f}_t(\boldsymbol{\theta}, \bar{f}_1) - \tilde{f}_t(\boldsymbol{\theta})| \to 0$ a.s.; see Straumann and Mikosch (2006). If $\{y_t\}_{t\in\mathbb{Z}}$ is SE, then the following log moment and contraction conditions ensure filter invertibility as well as e.a.s. convergence of the filtered parameter $\{\tilde{f}_t(\boldsymbol{\theta}, \bar{f}_1)\}_{t\in\mathbb{N}}$ to a strictly stationary and ergodic (SE) limit process $\{\tilde{f}_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ with

$$\mathbb{E}\log^+|\phi_y(f, y_t; \boldsymbol{\theta})| < \infty, \qquad \mathbb{E}\log\sup_f |\phi'_y(f, y_t; \boldsymbol{\theta})| < 0,$$

where $\phi'_y(f, y_t; \boldsymbol{\theta}) := \partial\phi_y(f, y_t; \boldsymbol{\theta})/\partial f$; see Straumann and Mikosch (2006) and Wintenberger (2013). Note that the expectations above are always taken w.r.t. the true measure, which may be indexed by $\boldsymbol{\theta}_0 \in \Theta$ when the model is correctly specified. Also note that we can establish the SE nature of the data $\{y_t\}_{t\in\mathbb{Z}}$ under correct specification by studying the ϕ_u recursion as in Section 2.1.

If the filtered sequence is analyzed at $\boldsymbol{\theta}_0$ and under correct initialization $\bar{f}_1 = f_1(\boldsymbol{\theta}_0)$ a.s., then we can clearly re-write the filtered parameter as

$$\tilde{f}_{t+1}(\boldsymbol{\theta}_0, \bar{f}_1) = \phi_u\big(\tilde{f}_t(\boldsymbol{\theta}_0, \bar{f}_1), u_t; \boldsymbol{\theta}_0\big) \quad \forall\, t \in \mathbb{N}, \tag{5$'$}$$

because $\{\tilde{f}_{t+1}(\boldsymbol{\theta}_0, \bar{f}_1)\}_{t\in\mathbb{N}} = \{f_t(\boldsymbol{\theta}_0)\}_{t\in\mathbb{N}}$ a.s. So without correct initialization, the contraction in ϕ_u does not ensure the invertibility of the filter nor its e.a.s. convergence to an SE limit, not even at $\boldsymbol{\theta}_0$. Invertibility is only ensured by the contraction in ϕ_y, which may differ considerably from that in ϕ_u since in general

$$\phi'_u(f, u_t; \boldsymbol{\theta}_0) \;=\; \partial\phi_y(f, g(f, u_t); \boldsymbol{\theta}_0)/\partial f \;\neq\; \partial\phi_y(f, y_t; \boldsymbol{\theta}_0)/\partial f \;=\; \phi'_y(f, y_t; \boldsymbol{\theta}_0).$$

The e.a.s. convergence of the filter $\{\tilde{f}_t(\boldsymbol{\theta}_0, \bar{f}_1)\}_{t\in\mathbb{N}}$ at $\boldsymbol{\theta}_0$ to an SE limit is implied by the contraction in ϕ_u, but only if one is willing to assume correct initialization. The contraction in ϕ_y also ensures the vanishing effect of the initial condition as it implies that the difference $|\tilde{f}_t(\boldsymbol{\theta}, \bar{f}_1) - \tilde{f}_t(\boldsymbol{\theta}_0)|$ vanishes asymptotically e.a.s. since

$$\big|\tilde{f}_{t+1}(\boldsymbol{\theta}, \bar{f}_1) - \tilde{f}_{t+1}(\boldsymbol{\theta})\big| = \big|\phi'_y\big(\tilde{f}^*_t(\boldsymbol{\theta}_0, \bar{f}_1), y_t; \boldsymbol{\theta}\big)\big| \times \big|\tilde{f}_t(\boldsymbol{\theta}, \bar{f}_1) - \tilde{f}_t(\boldsymbol{\theta})\big|.$$

2.3 Relations Between the DGP and the Filter

So far, we have seen that the contractions in ϕ_u and ϕ_y play very different roles as sufficient conditions for characterizing the observation-driven model as a data generating process or as a filter. Both ϕ_u and ϕ_y play unique roles irrespective of considering $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ or $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$. Proposition 1 further highlights this difference by showing that $\phi_u(f, u_t; \boldsymbol{\theta}) = \phi_y(f, y_t; \boldsymbol{\theta})$ holds only at $f = f_t$, and that $\phi'_u(f, u_t; \boldsymbol{\theta}) \neq \phi'_y(f, y_t; \boldsymbol{\theta})$ holds for any $f \in \mathcal{F} \subseteq \mathbb{R}$ in a large class of observation-driven models.

Proposition 1. *For every $f \in \mathcal{F}$ and $\boldsymbol{\theta} \in \Theta$, let the function $g \in \mathbb{C}^2(\mathcal{F} \times \mathcal{U})$ in (3) and $\phi_y \in \mathbb{C}$ in (4) satisfy the conditions*

$$\partial g(f,u)/\partial f \neq 0, \qquad \partial \phi_y(f,y;\boldsymbol{\theta})/\partial y \neq 0, \qquad \partial^2 \phi_y(f,y;\boldsymbol{\theta})/\partial f \partial y = 0, \qquad (6)$$

for almost every $(f,u,y) \in \mathcal{F} \times \mathcal{U} \times \mathcal{Y}$. Then the recursions ϕ_u and ϕ_y satisfy a.s.

$$\begin{aligned} \phi_u(f,u_t;\boldsymbol{\theta}) &\neq \phi_y(f,y_t;\boldsymbol{\theta}) \quad \forall\ (\boldsymbol{\theta},f) \in \Theta \times \mathcal{F} : f \neq f_t, \\ \phi_u'(f,u_t;\boldsymbol{\theta}) &\neq \phi_y'(f,y_t;\boldsymbol{\theta}) \quad \forall\ (\boldsymbol{\theta},f) \in \Theta \times \mathcal{F}. \end{aligned}$$

Proposition 1 highlights that the contraction conditions needed to establish model model invertibility and the e.a.s. convergence of the filtered parameter $\{\tilde{f}_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ to an SE limit are, for a large class of models, entirely different from those needed for establishing the SE nature of the true time-varying parameter $\{f_t(\boldsymbol{\theta}_0)\}_{t\in\mathbb{Z}}$ and the data $\{y_t\}_{t\in\mathbb{Z}}$. The first claim of Proposition 1 is obtained by noting that for every $\boldsymbol{\theta} \in \Theta$ and any $f \in \mathcal{F}$ we have

$$\phi_u(f,u_t;\boldsymbol{\theta}) = \phi_y(f,g(f,u_t);\boldsymbol{\theta}) = \phi_y(f,y_t^*;\boldsymbol{\theta}), \qquad (7)$$

with $y_t^* = y_t$ if and only if $f = f_t$. The second claim is obtained by noting that even at $f = f_t$ and $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ we have

$$\phi_u'(f_t,u_t;\boldsymbol{\theta}_0) = \phi_y'(f_t,y_t;\boldsymbol{\theta}_0) + e_y(f_t,y_t;\boldsymbol{\theta}_0), \qquad (8)$$

where $e_y(f_t,y_t;\boldsymbol{\theta}_0) := \big(\partial \phi_y(f_t,y_t;\boldsymbol{\theta}_0)/\partial y\big) \times \big(\partial g(f_t,u_t)/\partial f\big) \neq 0$ (a.s.) and evaluated at u_t satisfying $y_t = g(f_t,u_t)$. We call $e_y(f_t,y_t;\boldsymbol{\theta}_0)$ the *'feedback'* of the observation-driven model as it measures the impact of f_t on f_{t+1} via its impact on y_t. This is clear by re-writing the feedback as

$$e_y(f_t,y_t;\boldsymbol{\theta}_0) = (\partial f_{t+1}/\partial y_t) \ \times \ (\partial y_t/\partial f_t).$$

When convenient, we will re-write the feedback in terms of the innovations

$$e_u(f_t,u_t;\boldsymbol{\theta}_0) = e_y(f_t,g(f_t,u_t);\boldsymbol{\theta}_0) = e_y(f_t,y_t;\boldsymbol{\theta}_0).$$

Similarly to the relation between ϕ_u and ϕ_y, the relation between e_u and e_y is also subtle. In particular, it is important to keep in mind the inequality

$$e_u(f,u_t;\boldsymbol{\theta}_0) = e_y(f,g(f,u_t);\boldsymbol{\theta}_0) \neq e_y(f,y_t;\boldsymbol{\theta}_0)\ \forall\ f \neq f_t.$$

In Proposition 1, the first two inequalities in (6) are crucial to ensure that the time-varying parameter has an impact on the distribution of the data and that the data a.s. influences the path of the time-varying parameter. This makes the model 'observation-driven' as otherwise the observations would not necessarily drive the time-varying parameter, nor would the time-varying parameter necessarily drive the distribution of the data.

The zero cross-derivative condition restricts our attention to the class of observation driven models where the derivative ϕ_y' does not depend on the data. Many observation-driven models like the GARCH and the ACD model of Engle and Russell (1998) satisfy this property. In any case, this condition can easily be avoided by instead imposing the following a.s. restriction on the feedback,

$$\phi'_y(f, y_t^*; \boldsymbol{\theta}) - \phi'_y(f, y_t; \boldsymbol{\theta}) \neq -e(f, y_t^*; \boldsymbol{\theta}),$$

where $y_t^* := g(f, y_t; \boldsymbol{\theta})$.

Proposition 2 offers a counterpart to Proposition 1 by showing that despite the a.s. difference between $\phi_u(f, u_t; \boldsymbol{\theta})$ and $\phi_y(f, y_t; \boldsymbol{\theta})$ as well as the derivatives $\phi'_u(f, u_t; \boldsymbol{\theta})$ and $\phi'_y(f, y_t; \boldsymbol{\theta})$, there are still cases in which the contraction in ϕ_u implies (and is implied by) the contraction in ϕ_y. Blasques et al. (2014b) make implicit use of such conditions in stating some unproven local asymptotic results. As we shall see here, these conditions are quite restrictive and available only in specific cases.

PROPOSITION 2. *Let the observation-driven model defined in (3) and (4) satisfy*

$$\frac{\partial^2 \phi_y(f, y; \boldsymbol{\theta})}{\partial f \partial y} = \frac{\partial^2 \phi_y(f, y; \boldsymbol{\theta})}{\partial y^2} = 0 \quad \forall\ (f, y) \in \mathcal{F} \times \mathcal{Y} \subseteq \mathbb{R}^2, \tag{9}$$

for some $\boldsymbol{\theta} \in \Theta$. Then there exists a $\delta > 0$ such that if

$$\left|\mathbb{E} \log \sup_f |e_y(f, y_t; \boldsymbol{\theta})|\right| < \delta, \tag{10}$$

then

$$\mathbb{E} \log \sup_f |\phi'_u(f, u_t; \boldsymbol{\theta})| < 0 \quad \Leftrightarrow \quad \mathbb{E} \log \sup_f |\phi'_y(f, y_t; \boldsymbol{\theta})| < 0. \tag{11}$$

Condition (10) in Proposition 2 requires that the feedback effect is sufficiently small. This is a natural requirement for obtaining the contraction equivalence (11), since (8) shows that the only difference between the two contractions is precisely that ϕ_u takes the feedback into account, whereas ϕ_y does not. In essence, if there were no feedback effect, then both recursions would be the same. As pointed out in Proposition 1, however, the feedback cannot be zero as it is a fundamental feature of observation-driven models.

The second derivative conditions in (9) can be explained by the fact that (8) only holds at $f = f_t$. For every other $f \neq f_t$, the relation between ϕ_u and ϕ_y becomes more complex as shown in (7). These conditions hold easily in models where ϕ_y is a linear function of y_t. For example, the autoregressive conditional duration (ACD) model of Engle and Russell (1998) has

$$\phi_y(f_t, y_t; \boldsymbol{\theta}) = \omega + \alpha y_t + \beta f_t,$$

and hence (9) holds trivially. These conditions do not hold however in the case of the original GARCH model since the second derivative of ϕ_y w.r.t. y_t is α, and this parameter must satisfy the inequality $\alpha \neq 0$ for the model to satisfy the conditions of Proposition 1. In many cases, however, we can reformulate the measurement equation (3) such that the conditions in (9) easily hold again. For example, for the GARCH model we can equivalently define the whole model in terms of $z_t = y_t^2$ rather than in terms of y_t. We then obtain $\phi_y(f, z_t; \boldsymbol{\theta}) = \omega + \alpha z_t + \beta f$, such that the derivative condition is satisfied with respect to z_t rather than y_t.

Proposition 3 provides an alternative to Proposition 2 by relaxing the possibly problematic second-derivative condition in (9). This allows a direct application to the original GARCH model.

Proposition 3. *Let the observation-driven model defined in (3) and (4) satisfy,*

$$\frac{\partial^2 \phi_y(f, y; \boldsymbol{\theta})}{\partial f \partial y} = 0 \ \forall \ (f, y) \in \mathcal{F} \times \mathcal{Y} \subseteq \mathbb{R}^2, \tag{12}$$

for some $\boldsymbol{\theta} \in \Theta$. *Then there exists a* $\delta > 0$ *such that if*

$$\left| \mathbb{E} \log \sup_f |e_u(f, u_t; \boldsymbol{\theta})| \right| < \delta, \tag{13}$$

then

$$\mathbb{E} \log \sup_f |\phi'_u(f, u_t; \boldsymbol{\theta})| < 0 \quad \Leftrightarrow \quad \mathbb{E} \log \sup_f |\phi'_y(f, y_t; \boldsymbol{\theta})| < 0.$$

The use of e_u instead of e_y in Proposition 3 allows us to relax the second order derivative condition in (9) of Proposition 2. Consider again the GARCH model. Proposition 3 holds easily since

$$\frac{\partial^2 \phi_y(f, y; \boldsymbol{\theta})}{\partial f \partial y} = \frac{\partial \beta}{\partial y} = 0 \ \forall \ (f, y) \in \mathcal{F} \times \mathcal{Y},$$

and the feedback $e_u(f, u_t; \boldsymbol{\theta})$ is given by

$$e_u(f, u_t; \boldsymbol{\theta}) = 2\alpha f_t^{1/2} u_t \times \frac{1}{2} f_t^{-1/2} u_t = \alpha u_t^2 > 0 \text{ a.s.}$$

Hence, the uniform log feedback moment condition at $\boldsymbol{\theta} \in \Theta$ stated in (12) of Proposition 3 is given by

$$\mathbb{E} \log \sup_f |e_u(f, u_t; \boldsymbol{\theta})| \big| = \left| \mathbb{E} \log \sup_f |\alpha u_t^2| \right| = |\mathbb{E} \log |\alpha u_t^2|| < \delta.$$

The feedback at $\boldsymbol{\theta}$ thus becomes smaller and the contraction in ϕ_u approaches the contraction in ϕ_y at $\boldsymbol{\theta}$ as $\log |\alpha|$ gets close to $-\mathbb{E} \log |u_t^2|$.

Proposition 4 shows that the differentiability conditions can be completely avoided when dealing with uniform contraction conditions on ϕ_u and ϕ_y. As we shall see, the uniform contraction conditions can play an important role in obtaining bounded unconditional moments for the filtered process $\{\tilde{f}_t(\boldsymbol{\theta}, \bar{f}_1)\}_{t \in \mathbb{N}}$ as well as for the true $\{f_t(\boldsymbol{\theta}_0)\}_{t \in \mathbb{Z}}$ and the data $\{y_t\}_{t \in \mathbb{Z}}$ when working with an axiom of correct specification.

Proposition 4. *For any observation-driven model defined as in (3) and (4) and every* $\boldsymbol{\theta} \in \Theta$, *there exists a* $\delta > 0$ *such that if*

$$\sup_{f,y} |e_y(f, y; \boldsymbol{\theta})| < \delta, \quad \vee \quad \sup_{f,u} |e_u(f, u; \boldsymbol{\theta})| < \delta, \tag{14}$$

then

$$\begin{aligned} \mathbb{E} \log \sup_f |\phi'_u(f, u_t; \boldsymbol{\theta})| < 0 \quad &\Leftarrow \quad \sup_{f,u} |\phi'_u(f, u; \boldsymbol{\theta})| < 1 \\ \Leftrightarrow \quad \sup_{f,y} |\phi'_y(f, y; \boldsymbol{\theta})| < 1 \quad &\Rightarrow \quad \mathbb{E} \log \sup_f |\phi'_y(f, y_t; \boldsymbol{\theta})| < 0. \end{aligned}$$

In general, observation-driven models feature a feedback $e_y(f_t, y_t; \boldsymbol{\theta})$ that is either positive or negative almost surely. To make use of this property, we introduce the following definition.

Definition 1 (weak and strong stability). An observation-driven model is said to produce a *weakly stable* feedback at $\boldsymbol{\theta}$ if

$$\operatorname{sign}\big(e_y(f, y; \boldsymbol{\theta})\big) \neq \operatorname{sign}\big(\phi_y'(f, y^*; \boldsymbol{\theta})\big) \quad \wedge \quad |e_y(f, y; \boldsymbol{\theta})| < 2|\phi_y'(f, y^*; \boldsymbol{\theta})|, \tag{15}$$

for almost every $(f, y) \in \mathcal{F} \times \mathcal{Y}$ and $y = y^*$. The feedback is *strongly stable* if (15) holds for almost every $(f, y, y^*) \in \mathcal{F} \times \mathcal{Y}^2$. The feedback is *weakly unstable* if

$$\operatorname{sign}\big(e_y(f, y; \boldsymbol{\theta})\big) = \operatorname{sign}\big(\phi_y'(f, y^*; \boldsymbol{\theta})\big) \quad \vee \quad |e_y(f, y; \boldsymbol{\theta})| > 2|\phi_y'(f, y^*; \boldsymbol{\theta})|, \tag{16}$$

for almost every $(f, y) \in \mathcal{F} \times \mathcal{Y}$ and $y^* = y$. The feedback is *strongly unstable* if (16) holds for almost every $(f, y, y^*) \in \mathcal{F} \times \mathcal{Y}^2$.

For example, in the case of the GARCH model and for the empirically relevant case of $\alpha, \beta > 0$, we have $e_y(f, y; \boldsymbol{\theta}) = \alpha y^2/f > 0 \; \forall \; (f, y) \in \mathcal{F} \times \mathcal{Y}$ since $f \in \mathcal{F} \subseteq \mathbb{R}^+$. Since $\phi_y'(f, y; \boldsymbol{\theta}_0) = \beta > 0$, we directly obtain that the GARCH feedback is strongly unstable. The unstable nature of the GARCH feedback is not surprising: the model is precisely designed with the idea that a high (low) volatility at time t produces further high (low) volatilities at time $t + 1$.

In observation-driven models with unstable feedback mechanisms, the contraction in ϕ_u implies the contraction in ϕ_y as the former is more difficult to obtain than the latter. On the other hand, the contraction in ϕ_y does not necessarily imply the contraction in ϕ_u, since the latter still accounts for the feedback effect. As noted above, the GARCH model provides an immediate example since for $\alpha > 0$ and $\beta > 0$ we have

$$\mathbb{E} \log \sup_f |\phi_u'(f, u_t; \boldsymbol{\theta})| = \mathbb{E} \log |\alpha u_t^2 + \beta| < 0 \quad \Rightarrow \quad \mathbb{E} \log \sup_f |\phi_y'(f, y; \boldsymbol{\theta})| = \log |\beta| < 0.$$

Conversely, when the feedback effect is stable then the contraction in ϕ_y implies the contraction in ϕ_u as the latter takes the stabilizing feedback into account. Proposition 5 below highlights the intuitive role of the feedback effects e_y and e_u from Propositions 2 and 3 in linking the contraction in ϕ_u to the contraction in ϕ_y.

Proposition 5. *Let the observation-driven model defined in (3) and (4) satisfy the conditions in (9). If the feedback is weakly stable then it follows that*

$$\mathbb{E} \log \sup_f |\phi_y'(f, y_t; \boldsymbol{\theta})| < 0 \quad \Rightarrow \quad \mathbb{E} \log \sup_f |\phi_u'(f, u_t; \boldsymbol{\theta})| < 0.$$

If the feedback is weakly unstable then it follows that

$$\mathbb{E} \log \sup_f |\phi_y'(f, y_t; \boldsymbol{\theta})| < 0 \quad \Leftarrow \quad \mathbb{E} \log \sup_f |\phi_u'(f, u_t; \boldsymbol{\theta})| < 0.$$

Note that Proposition 5 allows the feedback effect to be large, as long as it is in the 'right direction'. Furthermore, Proposition 5 does not establish equivalence between the contraction in ϕ_u and ϕ_y. Instead, it describes the conditions under which one contraction implies another. This contrasts sharply with Propositions 2-4, which focus on contraction equivalence when the feedback effect is small.

Proposition 6 shows that the second derivative condition (9) can be avoided in the presence of strong rather than weak forms of feedback stability.

PROPOSITION 6. *Let the observation-driven model defined in (3) and (4) satisfy condition (12). If the feedback is strongly stable, then*

$$\mathbb{E}\log\sup_f |\phi_y'(f, y_t; \boldsymbol{\theta})| < 0 \quad \Rightarrow \quad \mathbb{E}\log\sup_f |\phi_u'(f, u_t; \boldsymbol{\theta})| < 0.$$

If the feedback is strongly unstable, then

$$\mathbb{E}\log\sup_f |\phi_y'(f, y_t; \boldsymbol{\theta})| < 0 \quad \Leftarrow \quad \mathbb{E}\log\sup_f |\phi_u'(f, u_t; \boldsymbol{\theta})| < 0.$$

Note that whereas Proposition 3 reveals the conditions under which the contraction in ϕ_y implies the contraction in ϕ_u, Proposition 6 explains that, in general, the contraction in ϕ_u implies the contraction in ϕ_y for unstable feedback mechanisms.

Inspection of the proof of Proposition 6 reveals also that condition (12) can be relaxed by requiring that the cross-derivative to be sufficiently small. This is stated in Corollary 1.

COROLLARY 1. *Suppose that (12) does not hold. Then there exists an $\epsilon > 0$ such that the claims of Proposition 6 are still true as long as*

$$\sup_{f,y}\left|\frac{\partial^2\phi_y(f, y; \boldsymbol{\theta})}{\partial f \partial y}\right| < \epsilon.$$

Our final Proposition 7 states a useful result for bounded feedback terms.

PROPOSITION 7. *Let the observation-driven model defined in (3) and (4) satisfy (12) and*

$$\sup_{f,y} |e_y(f, y; \boldsymbol{\theta})| < 1 \quad \vee \quad \mathbb{E}\log\sup_f |e_u(f, u_t; \boldsymbol{\theta})| < 0.$$

Then

$$\mathbb{E}\log\sup_f |\phi_y'(f, y_t; \boldsymbol{\theta})| < 0 \quad \Rightarrow \quad \mathbb{E}\log\sup_f |\phi_u'(f, u_t; \boldsymbol{\theta})| < 0.$$

2.4 Maximum Likelihood Estimation of Static Parameters

For any $\boldsymbol{\theta} \in \Theta$, the likelihood function implied by the observation-driven model in (3) takes the form

$$\ell_T(\boldsymbol{\theta}, \bar{f}_1) = \frac{1}{T}\sum_{t=2}^{T}\Big(\tilde{p}_t(\boldsymbol{\theta}, \bar{f}_1) + \log\tilde{g}_t'(\boldsymbol{\theta}, \bar{f}_1)\Big), \tag{17}$$

where $\tilde{g}_t(\boldsymbol{\theta}, \bar{f}_1) := \tilde{g}(\tilde{f}_t(\boldsymbol{\theta}, \bar{f}_1), y_t) := g^{-1}(\tilde{f}_t(\boldsymbol{\theta}, \bar{f}_1), y_t)$, where g^{-1} denotes the inverse of g with respect to its second argument, and where $\tilde{g}_t'(\boldsymbol{\theta}, \bar{f}_1) := \tilde{g}'(\tilde{f}_t(\boldsymbol{\theta}, \bar{f}_1), y_t) := \partial\tilde{g}(\tilde{f}_t(\boldsymbol{\theta}, \bar{f}_1), y)/\partial y|_{y=y_t}$, and $\tilde{p}_t(\boldsymbol{\theta}, \bar{f}_1) := \log p_u(\tilde{g}_t(\boldsymbol{\theta}, \bar{f}_1); \boldsymbol{\theta})$. Consistency and asymptotic normality of the maximum likelihood estimator (MLE)

$$\hat{\boldsymbol{\theta}}_T(\bar{f}_1) \in \arg\max_{\boldsymbol{\theta}\in\Theta} \ell_T(\boldsymbol{\theta}, \bar{f}_1) \tag{18}$$

are thus obtained by studying the stochastic properties of the random sequence $\{\ell_T(\boldsymbol{\theta}, \bar{f}_1)\}_{t\in\mathbb{N}}$, which are in turn defined by the properties of the stochastic sequences $\{\tilde{p}_t(\boldsymbol{\theta}, \bar{f}_1)\}_{t\in\mathbb{N}}$ and $\{\log \tilde{g}_t'(\boldsymbol{\theta}, \bar{f}_1)\}_{t\in\mathbb{N}}$ in (17). The stochastic properties of the likelihood function depend on those of the data $\{y_t\}_{t\in\mathbb{Z}}$ and the filtered sequence $\{\tilde{f}_t(\boldsymbol{\theta}, \bar{f}_1)\}_{t\in\mathbb{N}}$ for every $\boldsymbol{\theta} \in \Theta$.

As we have seen above, the ϕ_u recursions in (5) and (5′) are important at $\boldsymbol{\theta}_0$ for describing the properties of the data $\{y_t\}_{t\in\mathbb{Z}}$ when working under an axiom of correct specification. The ϕ_y recursions in (4) and (4′) are relevant at any $\boldsymbol{\theta} \in \Theta$ for establishing model invertibility and the e.a.s. convergence of the filter to an SE limit. In general, we are thus interested in ensuring that the ϕ_u contraction holds at $\boldsymbol{\theta}_0$ and the ϕ_y contraction holds over Θ. As noted above, Propositions 2-7 can be used to switch between contraction conditions on ϕ_u and ϕ_y as these two conditions are related in many ways and are sometimes even equivalent. For concreteness, we describe below how Propositions 2-7 can be used in different contexts that often determine the strategy adopted for establishing asymptotic results. A similar remark holds for the recent results in Harvey and Luati (2014) that concentrate on the verification of conditions on ϕ_u rather than ϕ_y.

Case 1 (*local asymptotics for well-specified models)*
When the feedback is stable then, by Propositions 2-7, we often require only the contraction of ϕ_y at $\boldsymbol{\theta}_0$ and continuity of ϕ_y' to ensure contraction of ϕ_u at $\boldsymbol{\theta}_0$ and of ϕ_y in a neighborhood of $\boldsymbol{\theta}_0$. Alternatively, when the feedback is unstable, then we just need the contraction in ϕ_u at $\boldsymbol{\theta}_0$ and the continuity of either ϕ_u' or ϕ_y'. If the feedback term is sufficiently small, then the contractions are equivalent and we can work with either condition regardless of the stable or unstable nature of the feedback.

Case 2 (*global asymptotics for well-specified models*)
When the feedback is stable for every $\boldsymbol{\theta} \in \Theta$ then we can often work exclusively with the contraction of ϕ_y on Θ. Alternatively, when the feedback is unstable on Θ, then we just need the contraction of ϕ_u to hold over the parameter space Θ. If the feedback term is sufficiently small uniformly on Θ, then the contractions are equivalent over Θ and we can work with either condition.

Case 3 (*asymptotics for mis-specified models*)
When we abandon the axiom of correct specification, then the properties of observed data $\{y_t\}_{t\in\mathbb{Z}}$ cannot be derived from the structure of the GAS model. Since the stochastic properties of $\{y_t\}_{t\in\mathbb{Z}}$ are in such cases known by assumption, we only need to study the filtered sequence $\{\tilde{f}_t(\boldsymbol{\theta}, \bar{f}_1)\}_{t\in\mathbb{N}}$ which is generated by the ϕ_y recursion for any $\boldsymbol{\theta} \in \Theta$. Theorems 2 and 4 in Blasques et al. (2014b) obtain the strong consistency and asymptotic normality of the MLE for a mis-specified GAS model w.r.t. to some pseudo-true parameter that minimizes the Kullback-Leibler divergence between the the true probability measure of $\{y_t\}_{t\in\mathbb{Z}}$ and the model implied measure. In any case, as we shall see, we can still appeal to Propositions 2-7 and make use of contraction conditions on the ϕ_u recursion of the misspecified model in order to show that the model is invertible and that the filter is asymptotically SE.

Though the results in this paper are applicable to a wider class of observation-driven models and settings, we now focus on obtaining local asymptotic results in the subclass of GAS models through conditions on ϕ_u. This allows us to provide proofs for the yet unsubstantiated claims in Blasques et al. (2014b) that made implicit use of the fact that, under appropriate conditions, a study of the ϕ_u recursion suffices for local results under correct specification.

3 Asymptotic Properties of MLE for GAS Models

3.1 Stationarity, ergodicity and moments of GAS processes

In the remainder of this paper we specify the GAS model for a d_y-dimensional series y_t as an observation-driven model where ϕ_y and ϕ_u are given by

$$\phi_y(f, y_t; \boldsymbol{\theta}) \quad = \quad \omega + \alpha \; s_y\big(f, y_t; \lambda\big) + \beta f, \tag{19}$$

$$\phi_u(f, u_t; \boldsymbol{\theta}) \quad = \quad \omega + \alpha \; s_u\big(f, u_t; \lambda\big) + \beta f, \tag{20}$$

where $s_u\big(f, u_t; \lambda\big) := s_y\big(f, g(f, u_t); \lambda\big)$, $\lambda \in \Lambda \subseteq \mathbb{R}^{d_\lambda}$ a parameter vector, and $\boldsymbol{\theta}^\top = (\omega, \alpha, \beta, \lambda^\top) \in \Theta \subseteq \mathbb{R}^{3+d_\lambda}$. We assume that the density $p_u(\,\cdot\,; \boldsymbol{\theta}) \equiv p_u(\,\cdot\,; \lambda)$ depends on λ only and has support $\mathcal{U}$ containing an open set for every λ. The link function $g : \mathcal{F} \times \mathcal{U} \to \mathcal{Y}$ is strictly increasing in its second argument with inverse $\tilde{g}(f, \,\cdot\,) = g^{-1}(f, \,\cdot\,)$ for almost every $f \in \mathcal{F}$. Throughout we assume that $\tilde{g} \in \mathbb{C}^{(2,0)}(\mathcal{F} \times \mathcal{Y})$, $\tilde{g}' \in \mathbb{C}^{(2,0)}(\mathcal{F} \times \mathcal{Y})$, $\tilde{p} \in \mathbb{C}^{(2,2)}(\mathcal{G} \times \Lambda)$, and $S \in \mathbb{C}^{(2,2)}(\mathcal{F} \times \Lambda)$, where $\mathcal{G} := \tilde{g}(\mathcal{Y}, \mathcal{F})$.

In order to establish the asymptotic properties of the filtered sequences $\{\tilde{f}_t(\boldsymbol{\theta}, \bar{f}_1)\}_{t\in\mathbb{N}}$ as well as of the true sequence $\{f_t(\boldsymbol{\theta}_0)\}_{t\in\mathbb{Z}}$ and the data $\{y_t\}_{t\in\mathbb{Z}}$ under correct specification, we define the random derivative function $s_u'(f, u_t; \lambda) := \partial s_u(f, u_t; \lambda)/\partial f$ and the kth power of the supremum

$$\rho_{u,t}^k(\boldsymbol{\theta}) := \sup_{f^* \in \mathcal{F}^*} |\beta + \alpha\, s_u'(f^*, u_t; \lambda)|^k,$$

with $\mathcal{F} \subseteq \mathcal{F}^* \subset \mathbb{R}$. Similarly, we define $s_y'(f, y_t; \lambda) := \partial s_y(f, y_t; \lambda)/\partial f$ and

$$\rho_{y,t}^k(\boldsymbol{\theta}) := \sup_{f^* \in \mathcal{F}^*} |\beta + \alpha\, s_y'(f^*, y_t; \lambda)|^k.$$

Proposition 8 uses the contraction equivalence results established in Section 2 to derive the stochastic properties of both true and filtered sequences in a neighborhood of $\boldsymbol{\theta}_0$. This is relevant for obtaining the local asymptotics of the MLE. Below we let $B_\epsilon(\boldsymbol{\theta}_0)$ denote an ϵ neighborhood of $\boldsymbol{\theta}_0$; i.e. $B_\epsilon(\boldsymbol{\theta}_0) := \{\boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| < \epsilon\}$.

Proposition 8. *Let $\{u_t\}_{t\in\mathbb{Z}}$ be an i.i.d. sequence, g be continuous and assume that $\exists\ \bar{f}_1 \in \mathcal{F} \subseteq \mathcal{F}^*$ for a convex $\mathcal{F}$ such that*

$$\mathbb{E} \log^+ |s_u(\bar{f}_1, u_1; \lambda_0)| < \infty, \qquad \mathbb{E} \log \rho_{u,1}^1(\boldsymbol{\theta}_0) < 0. \tag{21}$$

Then $\{f_t(\boldsymbol{\theta}_0)\}_{t\in\mathbb{Z}}$ and $\{y_t\}_{t\in\mathbb{Z}}$ are SE. If furthermore, s_y' is continuous and s_y satisfies

$$\mathbb{E} \log^+ |s_y(\bar{f}_1, y_t; \lambda_0)| < \infty, \qquad \frac{\partial^2 s_y(f, y; \boldsymbol{\theta}_0)}{\partial f \partial y} = 0 \;\; \forall\ (f, y) \in \mathcal{F} \times \mathcal{Y}, \tag{22}$$

then there exists $\delta > 0$ and $\epsilon > 0$ such that if

$$\Big|\mathbb{E} \log \sup_f |e_u(f, u_t; \boldsymbol{\theta}_0)|\Big| < \delta, \tag{23}$$

then $\{\tilde{f}_t(\boldsymbol{\theta}, \bar{f}_1)\}_{t\in\mathbb{N}}$ *converges e.a.s. to the unique SE solution* $\{\tilde{f}_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ *for every* $\boldsymbol{\theta} \in B_\epsilon(\boldsymbol{\theta}_0) \subseteq \Theta$ *with* $\{\tilde{f}_t(\boldsymbol{\theta}_0)\}_{t\in\mathbb{Z}} = \{f_t(\boldsymbol{\theta}_0)\}_{t\in\mathbb{Z}}$ *a.s.*

Proposition 8 is based on the conditions of Theorem 3.1 in Bougerol (1993) and Straumann and Mikosch (2006) and extends Proposition 1 of Blasques et al. (2014b) by considering also the filter properties and points $\boldsymbol{\theta}$ in a neighborhood of $\boldsymbol{\theta}_0$. This extension was made possible by the contraction equivalence results in Section 2.

Recall from Section 2 that the cross derivative condition in (22) can be easily avoided by imposing stricter conditions on the feedback term or reformulating the measurement equation of the GAS model in terms of transformations of the original observable variables. Similarly, the condition on the feedback term can be relaxed by considering further second derivative conditions. Here we have considered a local result based on the contraction equivalence of Proposition 3, but alternative conditions could have been used by building on the other propositions in Section 2.

Proposition 9 shows that the moments of the true and filtered sequences can also be bounded locally in a neighborhood of $\boldsymbol{\theta}_0$. For random variables X and Y we write $X \perp Y$ if X and Y are independent.

Proposition 9. *Let the conditions of Proposition 8 hold and let* $\exists\ n_f > 0$ *such that*

$$\mathbb{E}|s_u(\bar{f}_1, u_1; \lambda_0)|^{n_f} < \infty, \qquad \mathbb{E}\rho_{u,t}^{n_f}(\boldsymbol{\theta}_0) < 1. \tag{24}$$

Then $\mathbb{E}|f_t(\boldsymbol{\theta}_0)|^{n_f} < \infty$. *If furthermore,*

$$\mathbb{E}|s_y(\bar{f}_1, u_1; \lambda_0)|^{n_f} < \infty, \qquad \tilde{f}_t(\boldsymbol{\theta}, \bar{f}_1) \perp \rho_{y,t}^{n_f}(\boldsymbol{\theta})\ \forall\ (t, \bar{f}_1; \boldsymbol{\theta}) \in \mathbb{N} \times \mathcal{F} \times \Theta, \tag{25}$$

then there exists $\delta > 0$, $\epsilon > 0$, *and* $\Theta \supseteq B_\epsilon(\boldsymbol{\theta}_0)$ *such that if*

$$\mathbb{E}\sup_f |e_u(f, u_t; \boldsymbol{\theta}_0)|^{n_f} < \delta, \tag{26}$$

then $\sup_t \mathbb{E}|\tilde{f}_t(\boldsymbol{\theta}, \bar{f}_1)|^{n_f} < \infty$ *for every* $\boldsymbol{\theta} \in \Theta$.

Following Blasques et al. (2014b) we note that the orthogonality conditions in (25) can be avoided by using the stricter uniform contraction condition and feedback bound

$$\sup_{u\in\mathcal{U}} \rho_{u,t}^1(\boldsymbol{\theta}_0) < 1, \qquad \sup_{f,u} |e_u(f, u; \boldsymbol{\theta}_0)| < \delta.$$

This substitution is made possible through application of Proposition 4.

Corollary 2 extends Proposition 8 using the stability properties of the feedback effect to establish local results. A crucial difference with respect to Proposition 8 is thus that we no longer rely on contraction equivalence. As a result, the 'size' of the feedback term is not restricted to a small δ.

Corollary 2. *Let* $\{u_t\}_{t\in\mathbb{Z}}$ *be an i.i.d. sequence,* g *be continuous, and let* $\exists\ \bar{f}_1 \in \mathcal{F} \subseteq \mathcal{F}^*$ *for a convex* $\mathcal{F}$ *such that (21) holds. Suppose further that* s_y' *is continuous,* s_y *satisfies (22), and the feedback is strongly unstable at* $\boldsymbol{\theta}_0$. *Then* $\exists\ \epsilon > 0$ *and* $\Theta \supseteq B_\epsilon(\boldsymbol{\theta}_0)$ *such that* $\{\tilde{f}_t(\boldsymbol{\theta}, \bar{f}_1)\}_{t\in\mathbb{N}}$ *converges e.a.s. to the unique SE solution* $\{\tilde{f}_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ *for every* $\boldsymbol{\theta} \in \Theta$.

Corollary 3 extends the moment bounds of Proposition 9 by again allowing the 'size' of the feedback term not to be restricted.

Corollary 3. *Let the conditions of Corollary 2 hold and suppose that (24) and (25) hold. Then* $\exists\ \epsilon > 0$ *and* $\Theta \supseteq B_\epsilon(\boldsymbol{\theta}_0)$ *such that* $\sup_t \mathbb{E}|\tilde{f}_t(\boldsymbol{\theta}, \bar{f}_1)|^{n_f} < \infty$ *for every* $\boldsymbol{\theta} \in \Theta$.

Note again that we can produce many similar corollaries exploring the different combinations of conditions used in Propositions 5–7. For example, we could substitute the strong stability conditions by weak stability conditions by appealing to Proposition 6.

Propositions 8 and 9 and Corollaries 2 and 3 are important for our subsequent analysis. They establish the asymptotic behavior of the filtered sequence $\{\tilde{f}_t(\boldsymbol{\theta}, \bar{f}_1)\}_{t\in\mathbb{N}}$ in a neighborhood of $\boldsymbol{\theta}_0$ as well as of the true sequence $\{f_t(\boldsymbol{\theta}_0)\}_{t\in\mathbb{Z}}$ and the data $\{y_t\}_{t\in\mathbb{Z}}$. These three properties are key to establishing consistency and asymptotic normality in the remainder of this section.

3.2 Consistency and Asymptotic Normality

Consider the MLE $\hat{\boldsymbol{\theta}}_T(\bar{f}_1)$ defined in (18) with log likelihood function ℓ_T given in (17). To ensure that we have a sufficient number of moments for the likelihood function and its derivatives for consistency and asymptotic normality, we introduce the notion of moment preserving maps. This notion provides a direct link between the number of moments of the data (n_y) and the filtered time varying parameter (n_f) and the number of moments for the likelihood function and its derivatives. As such, it allows us to formulate low-level conditions for consistency and asymptotic normality; see Blasques et al. (2014b) for further details.

Definition 2. (Moment Preserving Maps)
Let $\boldsymbol{x}_t(\boldsymbol{\theta}) = (x_{1,t}, \ldots, x_{q,t}(\boldsymbol{\theta}))^\top$ *be a random variable for every* $\boldsymbol{\theta} \in \Theta$. *A function* $h : \mathbb{R}^q \times \Theta \to \mathbb{R}$ *is said to be* $\boldsymbol{n}/m$*-moment preserving, denoted as* $h(\cdot;\boldsymbol{\theta}) \in \mathbb{M}(\boldsymbol{n}, m)$, *if and only if* $\mathbb{E}|x_{i,t}(\boldsymbol{\theta})|^{n_i} < \infty$ *for* $\boldsymbol{n} = (n_1, \ldots, n_q)$ *and* $i = 1, \ldots, q$ *implies* $\mathbb{E}|h(\boldsymbol{x}_t(\boldsymbol{\theta}); \boldsymbol{\theta})|^m < \infty$.

Many familiar functions are $\boldsymbol{n}/m$ moment preserving, for example bounded functions, polynomial functions, and many more. Taking the GARCH model (2) as an example, $s(f_t, y_t; \lambda) = y_t^2 - f_t$, we have $s \in \mathbb{M}((n_f, n_y), m)$ with $m = \min(n_f, n_y/2)$. A catalogue of relevant functions and their moment preserving properties is provided in Lemma TA.6 of Blasques et al. (2014b).

We now introduce our assumptions for consistency and asymptotic normality of the MLE. Let $\mathcal{U}$ denote the support of $u_t(\lambda_0)$.

Assumption 1.

(i) $\exists n_u \geq 0$ *with* $\mathbb{E}|u_t|^{n_u} < \infty$;

(ii) $g \in \mathbb{M}(\boldsymbol{n}, n_y)$ *with* $\boldsymbol{n} := (n_f, n_u)$ *and* $n_y \geq 0$;

(iii) $p_y(y|f;\lambda) = p_y(y|f';\lambda')$ *holds for almost every* $y \in \mathcal{Y}$ *iff* $f = f'$ *and* $\lambda = \lambda'$.

Condition *(i)* of Assumption 1 ensures that the innovations have a common finite moment of order n_u. Condition *(ii)* imposes that g is moment preserving and ensures the existence of n_y moments if n_f moments of the time varying parameter and n_u moments of the innovations exist, respectively. Finally, condition *(iii)* ensures that the static model $f_t \equiv f$ is well-identified, such that we obtain a well-separated maximum of the likelihood.

Assumption 2. *$\exists\ \bar{f} \in \mathcal{F} \subseteq \mathcal{F}^*$ and $n_f > 0$ such that either*

(i) $\mathbb{E}|s_u(\bar{f}, u_1; \lambda_0)|^{n_f} < \infty$, $\mathbb{E}\rho_{u,t}^{n_f}(\boldsymbol{\theta}_0) < 1$, *and* $\tilde{f}_t(\boldsymbol{\theta}, \bar{f}_1) \perp \rho_{y,t}^{n_f}(\boldsymbol{\theta})\ \forall\ (t, \bar{f}_1; \boldsymbol{\theta})$;

or

(i′) $\sup_{u \in \mathcal{U}} |s_u(\bar{f}, u; \lambda_0)| = s_u(\bar{f}; \lambda_0) < \infty$ *and* $\sup_{f^* \in \mathcal{F}^*} |\partial s_u(f^*; \lambda_0)/\partial f| < 1$.

Furthermore, Θ is compact, and for almost every $y \in \mathcal{Y}$ we have $\alpha_0\, \partial s_y(f, y; \lambda_0)/\partial y \neq 0$ for every $f \in \mathcal{F}$.

Condition *(i)* or *(i′)* ensures that the true $\{f_t^u(\boldsymbol{\theta}_0)\}$ is SE and has n_f moments by application of Proposition 8. Together with condition *(ii)* in Assumption 1 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 Assumption 2, together with the assumption that $\alpha \neq 0$ ensures that the data $\{y_t(\boldsymbol{\theta}_0)\}$ entering the update equation (20) renders the filtered $\{\tilde{f}_t\}$ stochastic and non-degenerate.

Moment preservation is a natural requirement in the consistency and asymptotic normality proofs, as the likelihood and its derivatives are nonlinear functions of the original data y_t, the filtered time-varying parameter $\tilde{f}_t(\boldsymbol{\theta}, \bar{f}_1)$, and its derivatives with respect to $\boldsymbol{\theta}$. Let $n_{\log \tilde{g}'}$ and $n_{\tilde{p}}$ define the moment preserving orders of $\log \tilde{g}'$ and $\tilde{p}$, respectively; i.e. let $\log \tilde{g}' \in \mathbb{M}(\boldsymbol{n}, n_{\log \tilde{g}'})$ and $\tilde{p} \in \mathbb{M}(\boldsymbol{n}, n_{\tilde{p}})$ where $\boldsymbol{n} := (n_f, n_y)$. Note that by defining $\boldsymbol{n} := (n_f, n_y)$, we abstract from moment requirements with respect to λ. The later can be done without loss of generality, as λ is not stochastic.

Assumption 3. $\min\{n_{\log \tilde{g}'}, n_{\tilde{p}}\} \geq 1$.

Finally, Assumption 4 below explores the results of Sections 2 and 3 in order to obtain local asymptotic results for the filtered sequence without directly assuming the contraction of the ϕ_y recursion.

Assumption 4. *s_y' is continuous and s_y satisfies*

$$\mathbb{E}|s_y(\bar{f}_1, y_1; \lambda_0)|^{n_f} < \infty \quad \wedge \quad \partial^2 s_y(f, y; \boldsymbol{\theta}_0)/\partial f \partial y = 0\ \ \forall\ (f, y) \in \mathcal{F} \times \mathcal{Y}.$$

Theorem 1 derives the existence and identification of the ML estimator and establishes the strong consistency of the estimator as $T \to \infty$.

Theorem 1. (Consistency) *Suppose that $\{y_t\}_{t\in\mathbb{Z}}$ is generated by the GAS model in* (3) *and* (4) *and let Assumptions 1–4 hold. If the feedback effect is strongly unstable at $\boldsymbol{\theta}_0$, then there exists $\Theta \supseteq B_\epsilon(\boldsymbol{\theta}_0)$ with $\epsilon > 0$ such that $\hat{\boldsymbol{\theta}}_T(\bar{f}_1) \overset{a.s.}{\to} \boldsymbol{\theta}_0$ as $T \to \infty$.*

Theorem 1 exploits the results established in Corollary 3. Alternative conditions are possible as well. For example, we could use Proposition 9 such that instead of assuming that the feedback was strongly unstable, we could establish consistency of the MLE under an assumption that the feedback is small is expectation.

Corollary 4. *If Assumptions 1–4 hold, there exists $\delta, \epsilon > 0$ and $\Theta \supseteq B_\epsilon(\boldsymbol{\theta}_0)$ such that if $\mathbb{E}\sup_f |e_u(f, u_t; \boldsymbol{\theta}_0)|^{n_f} < \delta$, then $\hat{\boldsymbol{\theta}}_T(\bar{f}_1) \overset{a.s.}{\to} \boldsymbol{\theta}_0$ as $T \to \infty$.*

To establish asymptotic normality, we need additional assumptions. To formulate bounds on the moments of the likelihood and its derivatives, it is useful to introduce the following notation. Let

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

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, $s^{(\boldsymbol{k})} \in \mathbb{M}(\boldsymbol{n}, n_s^{(\boldsymbol{k})})$, 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. Again we suppress the third argument of s, the parameter λ, from the moment preserving requirements without loss of generality. 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 $n_s^{ff} := n_s^{(2,0,0)}$, $n_s^{\lambda} := n_s^{(0,0,1)}$, $n_s^{\lambda\lambda} := n_s^{(0,0,2)}$, $n_s^{f\lambda} := n_s^{(1,0,1)}$ etc. Similar definitions hold for the functions $\log \tilde{g}'$ and $\tilde{p}$. 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 $\tilde{f}(\boldsymbol{\theta}, \bar{f}_1)$ with respect to $\boldsymbol{\theta}$ and evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, respectively, where

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

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

Finally, we also define

$$n_{\ell'} = \min\left\{n_{\tilde{p}}^{\lambda}\ ,\ \frac{n_{\log\tilde{g}'}^{f} n_f^{(1)}}{n_{\log\tilde{g}'}^{f} + n_f^{(1)}}\ ,\ \frac{n_{\tilde{p}}^{f} n_f^{(1)}}{n_{\tilde{p}}^{f} + n_f^{(1)}}\right\},$$

and

$$\begin{aligned} n_{\ell''} \quad = \quad \min\Bigg\{ & n_{\tilde{p}}^{\lambda\lambda}\ ,\ \frac{n_{\tilde{p}}^{f\lambda} n_f^{(1)}}{n_{\tilde{p}}^{f\lambda} + n_f^{(1)}}\ ,\ \frac{n_{\tilde{p}}^{ff} n_f^{(1)}}{2n_{\tilde{p}}^{ff} + n_f^{(1)}}\ , \\ & \frac{n_{\tilde{p}}^{f} n_f^{(2)}}{n_{\tilde{p}}^{f} + n_f^{(2)}}\ ,\ \frac{n_{\log\tilde{g}'}^{f} n_f^{(2)}}{n_{\log\tilde{g}'}^{f} + n_f^{(2)}}\ ,\ \frac{n_{\log\tilde{g}'}^{ff} n_f^{(1)}}{2n_{\log\tilde{g}'}^{ff} + n_f^{(1)}}\Bigg\}. \end{aligned}$$

The required moment condition can now be stated as follows.

Assumption 5. *$n_{\ell'} \geq 2$ and $n_{\ell''} \geq 1$.*

Theorem 2 establishes the $T^{-1/2}$-convergence and asymptotic normality of the MLE $\hat{\boldsymbol{\theta}}_T(\bar{f}_1)$.

Theorem 2. (Asymptotic Normality) *Suppose that $\{y_t\}_{t\in\mathbb{Z}}$ is generated by the GAS model in (3) and (4) and let Assumptions 1–5 hold.*

If the feedback effect is strongly unstable at $\boldsymbol{\theta}_0$, then there exists $\Theta \supseteq B_\epsilon(\boldsymbol{\theta}_0)$ with $\boldsymbol{\theta}_0 \in \text{int}(\Theta)$ and $\epsilon > 0$ such that

$$\sqrt{T}(\hat{\boldsymbol{\theta}}_T(\bar{f}_1) - \boldsymbol{\theta}_0) \xrightarrow{d} \mathrm{N}\big(0, \mathcal{I}^{-1}(\boldsymbol{\theta}_0)\big) \quad \textit{as} \quad T \to \infty,$$

where $\mathcal{I}(\boldsymbol{\theta}_0) := \mathbb{E}\tilde{\ell}_t''(\boldsymbol{\theta}_0)$ is the Fisher information matrix and $\tilde{\ell}_t$ denotes the likelihood contribution of the tth observation.

4 Examples

In this section we discuss several examples to illustrate how we can easily apply the theory from Sections 2 and 3 to GAS models that are of empirical interest. In addition, the examples reveal how the correct specification axiom sometimes makes proofs much simpler, while the opposite is true in other cases.

4.1 GAS Models for Time Varying Location and Fat Tails

Consider the time varying location model $y_t = f_t + \lambda_1^{-1/2} u_t$ as in Harvey and Luati (2014). Assume that u_t has a standard Student's t distribution with λ_2 degrees of freedom, such that $\lambda_1, \lambda_2 > 0$. When the unit scaling function $S(f;\lambda) = 1$ is adopted, the scaled score that drives the update is given by

$$s_y(f_t, y_t; \lambda) = -\frac{\lambda_1(\lambda_2+1)(f_t - y_t)}{\lambda_1(f_t - y_t)^2 + \lambda_2} \quad \text{and} \quad s_u(f_t, u_t; \lambda) = -\frac{(\lambda_2+1)\lambda_1^{3/2} u_t}{\lambda_1^2 u_t^2 + \lambda_2}. \quad (27)$$

If we adopt instead the inverse information scaling $S(f_t;\lambda) = \lambda_1^{-1}(\lambda_2+1)^{-1}(\lambda_2+3)$, then the scaled score is given by

$$s_y(f_t, y_t; \lambda) = \frac{(1+3\lambda_2^{-1})(y_t - f_t)}{(1+\lambda_1 \cdot (y_t - f_t)^2/\lambda_2)} \quad , \quad s_u(f_t, u_t; \lambda) = \frac{\lambda_1^{-1/2}(1+3\lambda_2^{-1})u_t}{(1+u_t^2/\lambda_2)}. \quad (28)$$

In both scaling cases, the contraction condition for ϕ_u is trivially satisfied for $|\beta| < 1$ as $s_u(f_t, u_t; \lambda)$ does not depend on f_t. The contraction in ϕ_y is however more difficult to establish, regardless of the adopted scaling function.

Unfortunately, since the contraction in ϕ_u alone is not sufficient to obtain the asymptotic results for the MLE established in Section 3, we cannot claim the consistency and asymptotic normality of the MLE. Below, we explore the relation between the contraction in ϕ_u and the contraction ϕ_y established in Section 2 to obtain the desired asymptotic results for the MLE. We consider first the case where the feedback effect is sufficiently small for the two contractions to be equivalent. Next, we consider the case where the feedback is strongly unstable, so that the contraction in ϕ_u implies the contraction in ϕ_y.

Negligible Feedback and Contraction Equivalence

The feedback term in the unit-scaling model $S(f;\lambda) = 1$ is given by

$$e_y(f_t, y_t; \boldsymbol{\theta}) = \frac{\partial f_{t+1}}{\partial y_t}\frac{\partial y_t}{f_t} = \alpha \frac{\lambda_1(\lambda_2+1)\left(\lambda_2 - \lambda_1(f_t - y_t)^2\right)}{\left(\lambda_1(f_t - y_t)^2 + \lambda_2\right)^2}.$$

Since the feedback term is uniformly bounded by

$$\sup_{f,y} |e_y(f_t, y_t; \boldsymbol{\theta})| = |\alpha|\, \lambda_1\, (1 + \lambda_2^{-1}), \quad (29)$$

we can make the feedback arbitrarily small, uniformly in y and f, by letting $\lambda_1 \to 0$. By Proposition 4, the contractions are thus equivalent for small enough λ_1. As a result, the contraction in ϕ_u implied by $|\beta| < 1$ becomes sufficient for the consistency

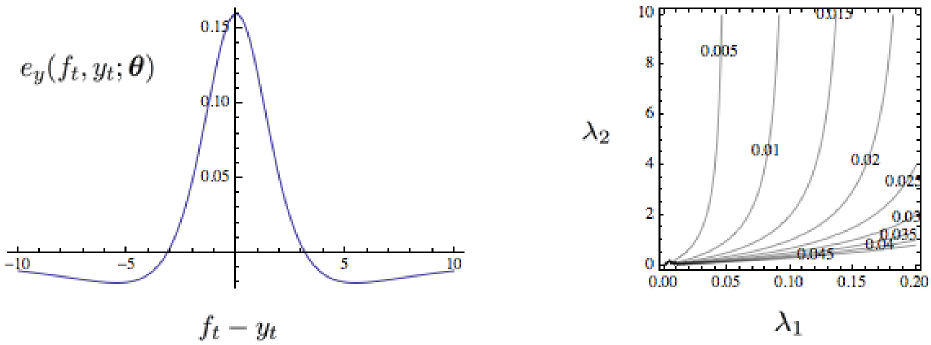

Figure 1: Value of feedback term $e_y(f_t, y_t; \boldsymbol{\theta})$ (left) as function of $f_t - y_t$ for $\alpha = 0.1$, $\lambda_1 = 0.5$ and $\lambda_2 = 5$ and contour plot (right) of maximum feedback value in equation (29)as function of λ_1 and λ_2.

and asymptotic normality of the MLE, as long as λ_1 is small enough. The restrictiveness of this condition depends on how close $|\beta|$ is to 1. Using the above conditions, the MLE for the GAS model is locally strongly consistent. Under additional moment conditions, we also obtain asymptotic normality.

Figure 1 plots the feedback $e_y(f_t, y_t; \boldsymbol{\theta})$ (left) as a function of $f_t - y_t$ for $\alpha = 0.1$, $\lambda_1 = 0.5$ and $\lambda_2 = 5$ as well as the contour plot (right) of its maximum value for different combinations of λ_1 and λ_2. The contour plot also reveals the pairs (λ_1, λ_2) for which the feedback is 'small enough', for any given value of $|\beta|$. For example, the region to the left of the contour line with value 0.02 indicates the admissible combinations of λ_1 and λ_2 for the case $|\beta| < 0.98$. Interestingly, the closer we get to the normal case (i.e. the larger the value of λ_2) the less restrictive the condition is on λ_1. This is the case considered in example 5.1 in Blasques et al. (2014). The claim made there that $|\beta| < 1$ is sufficient for obtaining asymptotic results holds true for the class of models with small enough λ_1.

Strong Feedback Instability and Model Invertibility

Consider now the case where the scaling function is the inverse conditional Fisher information matrix $S(f_t; \lambda) = \lambda_1^{-1}(\lambda_2 + 1)^{-1}(\lambda_2 + 3)$ with score given in (28). As noted above, we find ourselves once again in a setting where the contraction in ϕ_u is trivially implied by $|\beta| < 1$, but the contraction in ϕ_y is less obvious. We note that the cross-derivative statisfies

$$\frac{\partial^2 s_y(f_t, y_t; \lambda)}{\partial f_t \partial y_t} = -2\lambda_1\lambda_2^{-1}(1 + 3\lambda_2^{-1})\frac{(y_t - f_t)(\lambda_1\lambda_2^{-1}(y_t - f_t)^2 - 3)}{(1 + \lambda_1\lambda_2^{-1}(y_t - f_t)^2)^3}.$$

Its absolute value thus attains a maximum at $\lambda_1\lambda_2^{-1}(y_t - f_t)^2 = \sqrt{2} - 1$, such that

$$\left|\partial^2 s_y(f_t, y_t; \lambda)/\partial f_t \partial y_t\right| \leq 2\lambda_1^{1/2}\lambda_2^{-1/2}(1 + 3\lambda_2^{-1})\left(\frac{3}{8} + \frac{1}{4}\sqrt{2}\right).$$

We can make this arbitrarily small by taking the ratio λ_1/λ_2 close to zero. This is required by Corollary 1 if we wish to exploit the strong instability of the feedback effect. The effect on the non-linearity in equation (28) is clear. For small λ_1/λ_2 the non-linearity almost disappears, and the contraction condition for ϕ_y reduces in the limit of $\lambda_1/\lambda_2 = 0$ to $|\beta - \alpha(1+3\lambda_2^{-1})| < 1$. We can thus obtain the contraction in ϕ_y (and hence model invertibility) by appealing to Proposition 6, Corollary 1, and the fact that the feedback effect is strongly unstable under appropriate conditions. By Definition 1, this requires the sign of $\partial\phi_y(f_t, y_t; \boldsymbol{\theta})/\partial f_t$ to be equal to the sign of $e_y(f_t, y_t; \boldsymbol{\theta})$. For our model, we have

$$\frac{\partial\phi_y(f_t, y_t; \boldsymbol{\theta})}{\partial f_t} = \beta + \alpha\frac{\partial s_y(f_t, y_t; \lambda)}{\partial f_t} = \beta - \alpha\frac{\partial s_y(f_t, y_t; \lambda)}{\partial y_t} \tag{30}$$

$$= \beta - \frac{\partial\phi_y(f_t, y_t; \boldsymbol{\theta})}{\partial y_t} = \beta - e_y(f_t, y_t; \boldsymbol{\theta}). \tag{31}$$

Note that $s_y(f_t, y_t; \lambda)$ in (28) is uniformly bounded in f_t and y_t. We thus obtain strong instability if α is sufficiently small. Under these conditions, the MLE for the GAS model is locally strongly consistent. Under additional moment conditions, we also obtain asymptotic normality.

Consider the limiting case of the normal distribution, $\lambda_2^{-1} \to 0$. In that case, we get strong instability if $\mathrm{sign}(\beta - \alpha) = \mathrm{sign}(\alpha)$. As the contraction condition for ϕ_u is still $|\beta| < 1$, we obtain local consistency and asymptotic normality over the entire familiar triangle area $1 > \beta > \alpha > 0$. The regions can thus be quite wide.

4.2 GAS Models for Time Varying Scale and Fat Tails

Consider the GAS volatility model

$$y_t = g(f_t, u_t) = \lambda_1^{-1} f_t\, u_t, \quad u_t \sim p_u(u_t; \lambda), \tag{32}$$

with p_u denoting an $F(1, \lambda_2)$ distribution. Scaling by the inverse conditional Fisher information matrix $S(f_t; \lambda) = 2f_t^2 \cdot (1 + 3\lambda_2^{-1})$, we have

$$s_y(f_t, y_t; \lambda) = (1+3\lambda_2^{-1})\left(\frac{(1+\lambda_2^{-1})\,\lambda_1 y_t}{1+\lambda_1 y_t/(\lambda_2 f_t)} - f_t\right), \tag{33}$$

$$s_u(f_t, u_t; \lambda) = (1+3\lambda_2^{-1})\left(\frac{(1+\lambda_2^{-1})\,u_t}{1+u_t/\lambda_2} - 1\right)\cdot f_t, \tag{34}$$

$$\frac{\partial s_y(f_t, y_t; \lambda)}{\partial f_t} = (1+3\lambda_2^{-1})\left(\frac{(1+\lambda_2^{-1})\lambda_2^{-1}(\lambda_1 y_t/f_t)^2}{(1+\lambda_1 y_t/(\lambda_2 f_t))^2} - 1\right), \tag{35}$$

$$\frac{\partial^2 s_y(f_t, y_t; \lambda)}{\partial f_t \partial y_t} = 2(1+3\lambda_2^{-1})\left(\frac{(1+\lambda_2^{-1})\lambda_2^{-1}\lambda_1^2 y_t/f_t^2}{(1+\lambda_1 y_t/(\lambda_2 f_t))^3}\right). \tag{36}$$

To ensure positivity of the scale, we impose $\beta > (1+3\lambda_2^{-1})\alpha > 0$ and $\omega \geq \underline{\omega} > 0$.

This model embeds the Student's t GAS volatility model of Creal et al. (2011) and Harvey (2013) as a special case. If we let λ_2 diverge to infinity, we also recover the original GARCH model with normally distributed errors of Engle (1982) and Bollerslev (1986). This is easiliy seen by taking y_t to be the squared observations from the original GARCH model; see also Creal et al. (2013) and Harvey (2013) for details. For derivations of the asymptotic distribution properties of the MLE for GARCH models, we refer to the original contributions of Lee and Hansen (1994) and

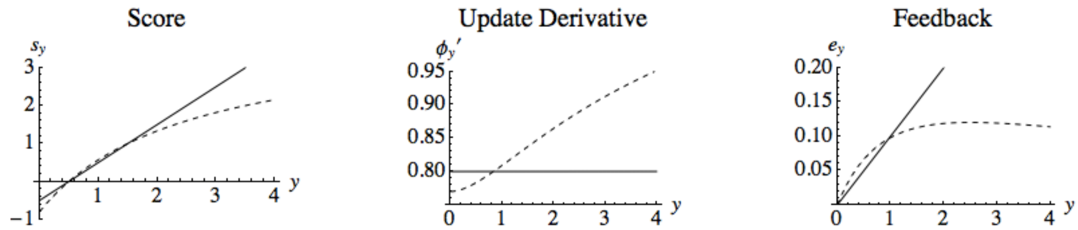

Figure 2: Score function s_y (left), derivative of update function ϕ'_y (center), and feedback effect $e_y(f_t, y_t; \boldsymbol{\theta})$ (right) for thin-tailed GAS model (solid) with $\lambda_2 \to \infty$ and fat tailed $\lambda_2 = 5$ GAS model (dashed) with $\lambda_1 = 1$, $\alpha = 0.1$ and $\beta = 0.8$.

Lumsdaine (1996), and references in the extensive reviews provided by Straumann (2005) and Francq and Zakoïan (2010).

Recall from Section 2 that the feedback effect in the GARCH model is strongly unstable and hence the contraction in ϕ_u implies the contraction in ϕ_y. As such, the invertibility condition $|\beta| < 1$ can be freely omitted as in Blasques et al. (2014b). The same argument applies to the fat tailed t-GARCH model. The fat tailed GAS model is however considerably more complicated to analyze due to its complex nonlinear dynamic behavior. The GAS scale model is also substantially different from the GAS location model considered in the previous section. Indeed, whereas in the location case s_u did not depend on f_t, now s_u is linear in f_t.

Figure 2 compares the score function s_y, the derivative of the update function ϕ'_y, and the feedback e_y of the thin-tailed and fat-tailed GAS models. In our scale GAS model, the contraction condition for $\phi_u(f_t, u_t; \boldsymbol{\theta})$ is given by

$$\mathbb{E} \log \left(\beta - \alpha \, (1 + 3\lambda_2^{-1}) + \alpha \, \frac{(1 + 3\lambda_2^{-1})(1 + \lambda_2^{-1}) \, u_t}{1 + u_t/\lambda_2} \right) < 0.$$

If we require a first order moment of f_t to exist, the log inside the expectations needs to be dropped, and we get the condition $1 > \beta > (1 + 3\lambda_2^{-1})\alpha > 0$; see also Blasques et al. (2012). The contraction in ϕ_y is however very difficult to ascertain. Figure 2 suggests however that the feedback effect is strongly unstable (since the feedback term e_y has the same sign as the update's derivative ϕ'_y) and hence that we might be able to make use of Proposition 6 and Corollary 1 to obtain the contraction in ϕ_y through the contraction in ϕ_u.

In order to apply Proposition 6 and Corollary 1, we note first from (36) that

$$0 \leq \frac{\partial^2 s_y(f_t, y_t; \lambda)}{\partial f_t \partial y_t} \leq \frac{8}{27}(1 + 3\lambda_2^{-1})(1 + \lambda_2^{-1})\lambda_1 \frac{1 - \beta}{\underline{\omega}},$$

as $f_t \geq \underline{\omega}/(1 - \beta)$ for all t if $f_1 > \underline{\omega}/(1 - \beta)$. We can thus make the cross derivative in (36) arbitrarily small if we set either λ_1 close to zero, β close to 1, or $\underline{\omega}$ sufficiently large. Alternatively, we can make the feedback effect small directly by picking α close to zero. Inspection of Figure 2 reveals that $\phi'_y(f_t, y_t; \boldsymbol{\theta}) > 0$ under the current restriction that $\beta > (1 + 3\lambda_2^{-1})\alpha > 0$ and that $\partial y_t / \partial f_t = \lambda_1^{-1} u_t > 0$. Whether or not the model is strongly unstable thus depends entirely on the sign of $\partial f_{t+1}/\partial y_t$. As

$$\partial f_{t+1}/\partial y_t = \alpha(1+3\lambda_2^{-1})(1+\lambda_2^{-1})\frac{\lambda_1}{(1+\lambda_1 y_t/(\lambda_2 f_t))^2} > 0,$$

almost surely, the model is strongly unstable as suggested by Figure 2. Using Proposition 6, Corollary 1 and Theorem 1, we thus obtain local strong consistency in the entire triangle region $\beta > (1+3\lambda_2^{-1})\alpha > 0$ and $\omega \geq \underline{\omega} > 0$. Under additional moment restrictions that shrink the region further, we also obtain asymptotic normality of the MLE. These results substantiate earlier unproven claims in Blasques et al. (2014b).

4.3 GAS count data model

Observation-driven models for time series of Poisson counts have been initiated by Davis et al. (2003). For the specific class of integer autoregression models, see Fokianos et al. (2009). In both contributions estimators and their asymptotic properties are developed. When we consider the GAS model for the Poisson distribution with intensity f_t and scaling by the conditional variance of the score, we obtain the general framework of Davis et al. (2003) with updating function for the Poisson intensity given by

$$f_{t+1} = \omega + \alpha(y_t - f_t) + \beta f_t, \tag{37}$$

with $\beta > \alpha > 0$. This recursion is linear in both y_t and f_t, such that $\mathbb{E}\tilde{\rho}_t^k(\boldsymbol{\theta}) = |\beta - \alpha| < 1$ is the key contraction condition for all k. The theory developed in this paper can thus also be used for establishing consistency and asymptotic normality for count data models under mis-specification. It appears that such results for GAS models have not been noted elsewhere. By contrast, adopting the correct specification axiom is problematic in this case: the recursion $\phi_u(f_t, u_t; \boldsymbol{\theta})$ is not continuous, which requires the use of alternative methods to the current contraction conditions to establish stationarity and ergodicity of the corresponding GAS data generated process. In contrast to the examples in Sections 4.1 and 4.2 where correct specification facilitated the contraction conditions on $\phi_u(f_t, u_t; \lambda)$, in the case of count data the correct specification axiom for GAS models seems to complicate matters considerably.

5 Final Remarks

To study consistency and asymptotic normality of the maximum likelihood estimator (MLE) of observation driven models under an axiom of correct specification typically requires the study of two non-linear dynamic systems: one for the *true*, and one for the *filtered* time varying parameter. In this paper we highlighted the intricate and subtle differences between the two sets of dynamic systems and the corresponding stability conditions. Moreover, we formulated a new set of restrictions on observation driven models under which the two sets of conditions are locally equivalent. If these restrictions hold, one only needs to verify one set of conditions to get local stability of both dynamic systems. In particular, we showed how these conditions could be used to establish model invertibility (for filtering purposes and consistency and asymptotic normality of the MLE), while only verifying stationarity and ergodicity of the true time varying parameter. We used these results to provide formal proofs of (as yet unsubstantiated) local asymptotics results for the MLE in generalized autoregressive score models. The results extended those in Harvey (2013) to score driven models that go beyond the MEM class and the exponential parameterization, and also help to clarify some of the subtle issues in the proofs of for example Harvey and Luati (2014) and Blasques et al. (2014b).

Appendix: Proofs

Proof of Proposition 1. If g is continuously differentiable in $\mathcal{F} \subseteq \mathbb{R}$ with strictly non-zero derivative, then it is also injective. Hence $g(f, y) = g(f', y)$ if and only if $f = f'$. With this in mind, the first claim follows immediately by noting that

$$\phi_u(f, u_t; \boldsymbol{\theta}) = \phi_y(f, g(f, u_t); \boldsymbol{\theta}) = \phi_y(f, y_t^*; \boldsymbol{\theta}) = \phi_y(f, y_t; \boldsymbol{\theta}) + \frac{\partial \phi_y(f, y_t^{**}; \boldsymbol{\theta})}{\partial y}(y_t^* - y_t),$$

and hence $\phi_u(f, u_t; \boldsymbol{\theta}) \neq \phi_y(f, y_t; \boldsymbol{\theta}) \quad \forall\ (\boldsymbol{\theta}, f) \in \Theta \times \mathcal{F} : f \neq f_t$ a.s. because $\partial \phi_y(f, y_t^{**}; \boldsymbol{\theta})/\partial y \neq 0$ a.s. $\ \forall\ f \in \mathcal{F}$, but $\phi_u(f, u_t; \boldsymbol{\theta}) = \phi_y(f, y_t; \boldsymbol{\theta})$ at $f = f_t$ a.s. because then $y_t^* = g(f, u_t; \boldsymbol{\theta}) = g(f_t, u_t; \boldsymbol{\theta}) = y_t$.

The second claim follows naturally by noting that

$$\begin{aligned}
\phi_u'(f, u_t; \boldsymbol{\theta}) &= \phi_y'(f, g(f, u_t); \boldsymbol{\theta}) + \frac{\partial \phi_y(f, g(f, u_t); \boldsymbol{\theta})}{\partial y} \frac{\partial g(f, u_t)}{\partial f} \\
&= \phi_y'(f, y_t; \boldsymbol{\theta}) + \frac{\partial^2 \phi_y(f, y_t^{**}; \boldsymbol{\theta})}{\partial f \partial y}(y_t^* - y_t) + \frac{\partial \phi_y(f, g(f, u_t); \boldsymbol{\theta})}{\partial y} \frac{\partial g(f, u_t)}{\partial f} \\
&= \phi_y'(f, y_t; \boldsymbol{\theta}) + \frac{\partial \phi_y(f, g(f, u_t); \boldsymbol{\theta})}{\partial y} \frac{\partial g(f, u_t)}{\partial f},
\end{aligned}$$

and hence $\phi_u'(f, u_t; \boldsymbol{\theta}) \neq \phi_y'(f, y_t; \boldsymbol{\theta})\ \forall\ f \in \mathcal{F}$ since by the condition

$$\frac{\partial \phi_y(f, g(f, u_t); \boldsymbol{\theta})}{\partial y} \frac{\partial g(f, u_t)}{\partial f} \neq 0 \quad \text{a.s.}$$

□

Proof of Proposition 2. The claim follows by noting that under the maintained assumptions

$$\begin{aligned}
\phi_u'(f, u_t; \boldsymbol{\theta}) &= \phi_y'(f, y_t; \boldsymbol{\theta}) + \frac{\partial^2 \phi_y(f, y_t^{**}; \boldsymbol{\theta})}{\partial f \partial y}(y_t^* - y_t) + e_y(f, y_t; \boldsymbol{\theta}) + \frac{\partial e_y(f, y_t^{***}; \boldsymbol{\theta})}{\partial y}(y_t^* - y_t) \\
&= \phi_y'(f, y_t; \boldsymbol{\theta}) + e_y(f, y_t; \boldsymbol{\theta}).
\end{aligned}$$

This implies that

$$\begin{aligned}
\mathbb{E} \log \sup_f |\phi_u'(f, u_t; \boldsymbol{\theta})| &= \mathbb{E} \log \sup_f |\phi_y'(f, u_t; \boldsymbol{\theta}) + e_y(f, y_t; \boldsymbol{\theta})| \\
&\leq \mathbb{E} \log \sup_f |\phi_y'(f, y_t; \boldsymbol{\theta})| + \mathbb{E} \log \sup_f |e_y(f, y_t; \boldsymbol{\theta})|,
\end{aligned}$$

and as a result

$$\left| \mathbb{E} \log \sup_f |\phi_u'(f, u_t; \boldsymbol{\theta})| - \mathbb{E} \log \sup_f |\phi_y'(f, y_t; \boldsymbol{\theta})| \right| \leq \left| \mathbb{E} \log \sup_f |e_y(f, y_t; \boldsymbol{\theta})| \right|.$$

□

Proof of Proposition 3. The claim follows by noting that under the maintained assumptions

$$\phi_u'(f, u_t; \boldsymbol{\theta}) = \phi_y'(f, y_t; \boldsymbol{\theta}) + \frac{\partial^2 \phi_y(f, y_t^{**}; \boldsymbol{\theta})}{\partial f \partial y}(y_t^* - y_t) + e_u(f, u_t; \boldsymbol{\theta}) = \phi_y'(f, y_t; \boldsymbol{\theta}) + e_u(f, u_t; \boldsymbol{\theta}),$$

which implies that

$$\begin{aligned}
\mathbb{E} \log \sup_f |\phi_u'(f, u_t; \boldsymbol{\theta})| &= \mathbb{E} \log \sup_f |\phi_y'(f, y_t; \boldsymbol{\theta}) + e_u(f, u_t; \boldsymbol{\theta})| \\
&\leq \mathbb{E} \log \sup_f |\phi_y'(f, y_t; \boldsymbol{\theta})| + \mathbb{E} \log \sup_f |e_u(f, u_t; \boldsymbol{\theta})|
\end{aligned}$$

$$\Rightarrow \quad \Big|\mathbb{E}\log\sup_f|\phi_u'(f,u_t;\boldsymbol{\theta})| - \mathbb{E}\log\sup_f|\phi_y'(f,y_t;\boldsymbol{\theta})|\Big| \leq \big|\mathbb{E}\log\sup_f|e_u(f,u_t;\boldsymbol{\theta})|\big|.$$

□

Proof of Proposition 4. The claim follows by noting that

$$\phi_u'(f,u_t;\boldsymbol{\theta}) = \phi_y'(f,y_t^*;\boldsymbol{\theta}) + e_y(f,y_t^*;\boldsymbol{\theta}) = \phi_y'(f,y_t^*;\boldsymbol{\theta}) + e_u(f,u_t;\boldsymbol{\theta},$$

with $y_t^* = g(f,u_t)$. Hence, $|\sup_{f,u}|\phi_u'(f,u;\boldsymbol{\theta})| - \sup_{f,y}|\phi_y'(f,y;\boldsymbol{\theta})|| \leq \big|\sup_{f,u}|e_u(f,u;\boldsymbol{\theta})|\big|$, and

$$\sup_{f,u}|\phi_u'(f,u;\boldsymbol{\theta})| = \sup_{f,u}|\phi_y'(f,g(f,u);\boldsymbol{\theta}) + e_y(f,g(f,u);\boldsymbol{\theta})| \;\leq \sup_{f,y}|\phi_y'(f,y;\boldsymbol{\theta})| + \sup_{f,y}|e_y(f,y;\boldsymbol{\theta})|$$

$$\Rightarrow \quad \Big|\sup_{f,u}|\phi_u'(f,u;\boldsymbol{\theta})| - \sup_{f,y}|\phi_y'(f,y;\boldsymbol{\theta})|\Big| \leq \big|\sup_f|e_y(f,y;\boldsymbol{\theta})|\big|.$$

□

Proof of Proposition 5. In the proof of Proposition 2 we showed that under the second-derivative conditions in (9) we have

$$\phi_u'(f,u_t;\boldsymbol{\theta}) = \phi_y'(f,y_t;\boldsymbol{\theta}) + e_y(f,y_t;\boldsymbol{\theta}),$$

and hence that

$$\mathbb{E}\log\sup_f|\phi_u'(f,u_t;\boldsymbol{\theta}_0)| = \mathbb{E}\log\sup_f|\phi_y'(f,y_t;\boldsymbol{\theta}_0) + e_y(f_t,y_t;\boldsymbol{\theta}_0)|.$$

Now, if the feedback is weakly stable then we have a.s.

$$\text{sign}\big(e_y(f,y_t;\boldsymbol{\theta})\big) \neq \text{sign}\big(\phi_y'(f,y_t;\boldsymbol{\theta})\big) \quad \wedge \quad |e_y(f,y_t;\boldsymbol{\theta})| < 2|\phi_y'(f,y_t;\boldsymbol{\theta})|,$$

which implies that $|\phi_y'(f,y_t;\boldsymbol{\theta}_0)| \geq |\phi_y'(f,y_t;\boldsymbol{\theta}_0) + e_y(f,y_t;\boldsymbol{\theta}_0)|$ holds a.s. and hence that

$$\log\sup_f|\phi_y'(f,y_t;\boldsymbol{\theta}_0)| \geq \log\sup_f|\phi_y'(f,y_t;\boldsymbol{\theta}_0) + e_y(f,y_t;\boldsymbol{\theta}_0)| = \mathbb{E}\log\sup_f|\phi_u'(f,u_t;\boldsymbol{\theta}_0)| \quad \text{a.s.}$$

As a result $\mathbb{E}\log\sup_f|\phi_y'(f,y_t;\boldsymbol{\theta}_0)| < 0 \Rightarrow \mathbb{E}\log\sup_f|\phi_u'(f,u_t;\boldsymbol{\theta}_0)| < 0$. Alternatively, if the feedback is weakly unstable then

$$\text{sign}\big(e_y(f,y_t;\boldsymbol{\theta})\big) = \text{sign}\big(\phi_y'(f,y_t;\boldsymbol{\theta})\big) \quad \vee \quad |e_y(f,y_t;\boldsymbol{\theta})| > 2|\phi_y'(f,y_t;\boldsymbol{\theta})| \quad \text{a.s.},$$

which implies that $|\phi_y'(f,y_t;\boldsymbol{\theta}_0)| \leq |\phi_y'(f,y_t;\boldsymbol{\theta}_0) + e_y(f,y_t;\boldsymbol{\theta}_0)|$ holds a.s. and hence

$$\mathbb{E}\log\sup_f|\phi_y'(f,y_t;\boldsymbol{\theta}_0)| \leq \mathbb{E}\log\sup_f|\phi_y'(f,y_t;\boldsymbol{\theta}_0) + e_y(f,y_t;\boldsymbol{\theta}_0)| = \mathbb{E}\log\sup_f|\phi_u'(f,u_t;\boldsymbol{\theta}_0)|,$$

and as a result $\mathbb{E}\log\sup_f|\phi_y'(f,y_t;\boldsymbol{\theta}_0)| < 0 \quad \Leftarrow \quad \mathbb{E}\log\sup_f|\phi_u'(f,u_t;\boldsymbol{\theta}_0)| < 0$. □

Proof of Proposition 6. The proof follows by noting that under (12)

$$\begin{aligned}\phi_u'(f,u_t;\boldsymbol{\theta}) &= \phi_y'(f,y_t^*;\boldsymbol{\theta}) + e_y(f,y_t^*;\boldsymbol{\theta}) = \phi_y'(f,y_t;\boldsymbol{\theta}) + \frac{\partial\phi_y'(f,y_t^{**};\boldsymbol{\theta})}{\partial y}(y_t^* - y_t) + e_y(f,y_t^*;\boldsymbol{\theta}) \\ &= \phi_y'(f,y_t;\boldsymbol{\theta}) + e_y(f,y_t^*;\boldsymbol{\theta}). \end{aligned} \tag{A1}$$

Hence, if the feedback is strongly stable, we have a.s.

$$\text{sign}\big(e_y(f,y_t^*;\boldsymbol{\theta})\big) \neq \text{sign}\big(\phi_y'(f,y_t;\boldsymbol{\theta})\big) \quad \wedge \quad |e_y(f,y_t^*;\boldsymbol{\theta})| < 2|\phi_y'(f,y_t;\boldsymbol{\theta})|,$$

which implies that $|\phi_y'(f,y_t;\boldsymbol{\theta})| \geq |\phi_y'(f,y_t;\boldsymbol{\theta}) + e_y(f,y_t^*;\boldsymbol{\theta})|$ holds a.s. and hence that

$$\mathbb{E}\log\sup_f|\phi_y'(f,y_t;\boldsymbol{\theta})| \geq \mathbb{E}\log\sup_f|\phi_y'(f,y_t;\boldsymbol{\theta}) + e_y(f,y_t^*;\boldsymbol{\theta})| = \mathbb{E}\log\sup_f|\phi_u'(f,u_t;\boldsymbol{\theta})|.$$

As a result $\mathbb{E}\log\sup_f|\phi_y'(f,y_t;\boldsymbol{\theta})| < 0 \Rightarrow \mathbb{E}\log\sup_f|\phi_u'(f,u_t;\boldsymbol{\theta}_0)| < 0$. Alternatively, if the feedback is strongly unstable then

$$\text{sign}\big(e_y(f,y_t^*;\boldsymbol{\theta})\big) = \text{sign}\big(\phi_y'(f,y_t;\boldsymbol{\theta})\big) \quad \vee \quad |e_y(f,y_t^*;\boldsymbol{\theta})| > 2|\phi_y'(f,y_t;\boldsymbol{\theta})| \quad \text{a.s.},$$

which implies that $|\phi_y'(f,y_t;\boldsymbol{\theta}_0)| \leq |\phi_y'(f,y_t;\boldsymbol{\theta})+e_y(f,y_t^*;\boldsymbol{\theta})|$ holds a.s. and hence

$$\mathbb{E}\log\sup_f|\phi_y'(f,y_t;\boldsymbol{\theta})| \leq \mathbb{E}\log\sup_f|\phi_y'(f,y_t;\boldsymbol{\theta})+e_y(f,y_t^*;\boldsymbol{\theta})| = \mathbb{E}\log\sup_f|\phi_u'(f,u_t;\boldsymbol{\theta})|,$$

and as a result $\mathbb{E}\log\sup_f|\phi_y'(f,y_t;\boldsymbol{\theta})|<0 \quad\Leftarrow\quad \mathbb{E}\log\sup_f|\phi_u'(f,u_t;\boldsymbol{\theta})|<0.$ □

Proof of Proposition 7. Using (A1), we have that

$$\begin{aligned}\mathbb{E}\log\sup_f|\phi_u'(f,u_t;\boldsymbol{\theta})| &= \mathbb{E}\log\sup_f|\phi_y'(f,y_t;\boldsymbol{\theta})+e_y(f,y_t^*;\boldsymbol{\theta})|\\ &\leq \mathbb{E}\log\sup_f|\phi_y'(f,y_t;\boldsymbol{\theta})|+\mathbb{E}\log\sup_f|e_y(f,y_t^*;\boldsymbol{\theta})|\\ &\leq \mathbb{E}\log\sup_f|\phi_y'(f,y_t;\boldsymbol{\theta})|+\log\sup_{f,y}|e_y(f,y;\boldsymbol{\theta})|,\end{aligned}$$

with $\log\sup_{f,y}|e_y(f,y;\boldsymbol{\theta})|<0$ since $\sup_{f,y}|e_y(f,y;\boldsymbol{\theta})|<1$, and thus

$$\mathbb{E}\log\sup_f|\phi_y'(f,y_t;\boldsymbol{\theta})|<0 \quad\Rightarrow\quad \mathbb{E}\log\sup_f|\phi_u'(f,u_t;\boldsymbol{\theta})|<0.$$

Alternatively, using the fact that

$$\phi_u'(f,u_t;\boldsymbol{\theta}) = \phi_y'(f,y_t;\boldsymbol{\theta})+\frac{\partial\phi_y'(f,y_t^{**};\boldsymbol{\theta})}{\partial y}(y_t^*-y_t)+e_y(f,y_t^*;\boldsymbol{\theta}) = \phi_y'(f,y_t;\boldsymbol{\theta})+e_u(f,u_t;\boldsymbol{\theta}),$$

we obtain

$$\begin{aligned}\mathbb{E}\log\sup_f|\phi_u'(f,u_t;\boldsymbol{\theta})| &= \mathbb{E}\log\sup_f|\phi_y'(f,y_t;\boldsymbol{\theta})+e_u(f,u_t;\boldsymbol{\theta})|\\ &\leq \mathbb{E}\log\sup_f|\phi_y'(f,y_t;\boldsymbol{\theta})|+\mathbb{E}\log\sup_f|e_u(f,u_t;\boldsymbol{\theta})|.\end{aligned}$$

Since $\mathbb{E}\log\sup_f|e_u(f,u_t;\boldsymbol{\theta})|<0$, it follows that

$$\mathbb{E}\log\sup_f|\phi_y'(f,y_t;\boldsymbol{\theta})|<0 \quad\Rightarrow\quad \mathbb{E}\log\sup_f|\phi_u'(f,u_t;\boldsymbol{\theta})|<0.$$

□

Proof of Proposition 8. Proposition 1 in Blasques et al. (2014b) shows that $\{f_t(\boldsymbol{\theta}_0)\}_{t\in\mathbb{Z}}$ is SE under the conditions in (21). By continuity of g and the iid nature of $\{u_t\}_{t\in\mathbb{Z}}$, we conclude that $\{y_t\}_{t\in\mathbb{Z}}$ is also SE by Krengel (1985). Proposition 3 shows that under condition (22), the contraction condition in (21) implies

$$\mathbb{E}\log\sup_f|\alpha_0 s_y'(f,y_t,\lambda_0)+\beta_0|<0.$$

By Proposition 2 in Blasques et al. (2014b) the filtered process $\{\tilde{f}_t(\boldsymbol{\theta}_0,\bar{f}_1)\}_{t\in\mathbb{N}}$ converges e.a.s. to the unique SE solution $\{\tilde{f}_t(\boldsymbol{\theta}_0)\}_{t\in\mathbb{Z}}$. Continuity of s_y and s_y' ensures that $\exists\ \epsilon>0$ and $\Theta\supseteq B_\epsilon(\boldsymbol{\theta}_0)$ such that $\{\tilde{f}_t(\boldsymbol{\theta},\bar{f}_1)\}_{t\in\mathbb{N}}$ converges e.a.s. to the unique SE solution $\{\tilde{f}_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ for every $\boldsymbol{\theta}\in\Theta$. Berge's maximum theorem ensures continuity of the supremum. □

Proof of Proposition 9. By Proposition 8, $\{f_t(\boldsymbol{\theta}_0)\}_{t\in\mathbb{Z}}$. Under the conditions in (24), Proposition 1 Blasques et al. (2014b) shows that $\mathbb{E}|f_t(\boldsymbol{\theta}_0)|^{n_f}<\infty$. A simple adaptation of the proof of Proposition 3 shows that under the cross-derivative condition in (22) and the feedback condition (26) we have that $\mathbb{E}\rho_{u,t}^{n_f}(\boldsymbol{\theta}_0)<1 \;\Rightarrow\; \mathbb{E}\rho_{y,t}^{n_f}(\boldsymbol{\theta}_0)<1$, because

$$\phi_u'(f,u_t;\boldsymbol{\theta}) = \phi_y'(f,y_t;\boldsymbol{\theta})+\frac{\partial^2\phi_y(f,y_t^{**};\boldsymbol{\theta})}{\partial f\partial y}(y_t^*-y_t)+e_u(f,u_t;\boldsymbol{\theta}) = \phi_y'(f,y_t;\boldsymbol{\theta})+e_u(f,u_t;\boldsymbol{\theta}),$$

which implies that

$$\begin{aligned}\mathbb{E}\sup_f|\phi_u'(f,u_t;\boldsymbol{\theta})|^{n_f} &= \mathbb{E}\sup_f|\phi_y'(f,y_t;\boldsymbol{\theta})+e_u(f,u_t;\boldsymbol{\theta})|^{n_f}\\ &\leq \mathbb{E}\sup_f|\phi_y'(f,y_t;\boldsymbol{\theta})|^{n_f}+\mathbb{E}\sup_f|e_u(f,u_t;\boldsymbol{\theta})|^{n_f}.[2]\end{aligned}$$

By Proposition 2 of Blasques et al. (2014b) and under condition (25), we then have $\sup_t \mathbb{E}|\tilde{f}_t(\boldsymbol{\theta}_0, \bar{f}_1)|^{n_f} < \infty$. Continuity of s_y and s_y' ensures that the same conditions hold in a neighborhood of $\boldsymbol{\theta}_0$ and hence that $\exists \epsilon > 0$ and $\Theta \supseteq B_\epsilon(\boldsymbol{\theta}_0)$ such that $\sup_t \mathbb{E}|\tilde{f}_t(\boldsymbol{\theta}, \bar{f}_1)|^{n_f} < \infty$ holds for every $\boldsymbol{\theta} \in \Theta$. Again, Berge's maximum theorem ensures continuity of the supremum. □

Proof of Theorem 1. Assumptions 2 and 4 imply the conditions of Proposition 3, thus establishing the equivalence of the contraction condition for $\phi_u'(f, u; \boldsymbol{\theta})$ in Assumption 2 to that of $\phi_y'(f, y; \boldsymbol{\theta})$ over some small compact neighborhood $\Theta \subseteq B_\epsilon(\boldsymbol{\theta}_0)$ for some small $\epsilon > 0$. Given that the contraction condition is satisfied for $\phi_y'(f, y; \boldsymbol{\theta})$ over this neighborhood, the rest of the proof follows directly along the same lines as the proof of Theorem 2 in Blasques et al. (2014b). See technical appendix for further details. □

Proof of Theorem 2. As in the proof of Theorem 1, the equivalence of the contraction condition for $\phi_u'(f, u; \boldsymbol{\theta})$ in Assumption 2 to that of $\phi_y'(f, y; \boldsymbol{\theta})$ over some small compact neighborhood $\Theta \subseteq B_\epsilon(\boldsymbol{\theta}_0)$ for some small $\epsilon > 0$ again follows from Proposition 3. Using the additional moment conditions in Assumption 5, the rest of the proof follows directly along the lines of the proof of Theorem 4 in Blasques et al. (2014b). See technical appendix for further details. □

References

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

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

Blasques, F., Koopman, S. J., and Lucas, A. (2012). Stationarity and ergodicity of univariate generalized autoregressive score processes. *Discussion Paper Tinbergen Institute TI 12-059/4.*

Blasques, F., Koopman, S. J., and Lucas, A. (2014a). Information theoretic optimality of observation driven time series models. *Discussion Paper Tinbergen Institute TI 14-046/III.*

Blasques, F., Koopman, S. J., and Lucas, A. (2014b). Maximum likelihood estimation for generalized autoregressive score models. *Discussion Paper Tinbergen Institute TI 14-029/III.*

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.

Cline, D. and Pu, H. (1999). Geometric ergodicity of nonlinear time series. *Statistica Sinica*, 9(4):1103–1118.

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

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

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

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

Davis, R., Dunsmuir, W., and Streett, S. (2003). Observation-driven models for poisson counts. *Biometrika*, 90:777–790.

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

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

Fokianos, K., Rahbek, A., and Tjostheim, D. (2009). Poisson autoregression. *Journal of the American Statistical Association*, 104:1430–1439.

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

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

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

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

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

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

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

Meyn, S. P. and Tweedie, R. L. (2009). *Markov chains and stochastic stability*. Cambridge University Press.

Ranga Rao, R. (1962). Relations between weak and uniform convergence of measures with applications. *Annals of Mathematical Statistics*, 33:659–680.

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

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

Van der Vaart, A. W. (2000). *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). Continuous invertibility and stable QML estimation of the EGARCH(1, 1) model. *Scandinavian Journal of Statistics*, 40(4):846–867.

Technical Appendix

Feedback Effects, Contraction Conditions and Asymptotic Properties of the Maximum Likelihood Estimator for Observation-Driven Models‡

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

[(a)] *VU University Amsterdam and Tinbergen Institute*
[(b)] *CREATES, Aarhus University*

Proof of Theorem 1. Existence of the ML estimator and identification of $\boldsymbol{\theta}_0$ is ensured under the present conditions by Theorems 1 and 3 in Blasques et al. (2014b). As a result, $\hat{\boldsymbol{\theta}}_T(\bar{f}_1)$ is a measurable map $\forall\, \bar{f}_1 \in \mathcal{F}$ and $\ell_\infty(\boldsymbol{\theta}_0) > \ell_\infty(\boldsymbol{\theta})\ \forall\, \boldsymbol{\theta} \in \Theta : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$. Following the classical consistency argument found e.g. in White (1994, Theorem 3.4) or Gallant and White (1988, Theorem 3.3), we obtain $\hat{\boldsymbol{\theta}}_T(\bar{f}_1) \xrightarrow{a.s.} \boldsymbol{\theta}_0$ from the uniform convergence of the criterion function and the identifiable uniqueness of the maximizer $\boldsymbol{\theta}_0 \in \Theta$

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

Step 1, uniform convergence: Let $\ell_T(\boldsymbol{\theta})$ denote the likelihood function $\ell_T(\boldsymbol{\theta}, \bar{f}_1)$ with $\tilde{f}_t(y^{1:t-1}, \boldsymbol{\theta}, \bar{f}_1)$ replaced by $\tilde{f}_t(y^{t-1}, \boldsymbol{\theta})$. 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 and note that

$$\sup_{\boldsymbol{\theta}\in\Theta} |\ell_T(\boldsymbol{\theta}, \bar{f}_1) - \ell_\infty(\boldsymbol{\theta})| \ \leq \sup_{\boldsymbol{\theta}\in\Theta} |\ell_T(\boldsymbol{\theta}, \bar{f}_1) - \ell_T(\boldsymbol{\theta})| + \sup_{\boldsymbol{\theta}\in\Theta} |\ell_T(\boldsymbol{\theta}) - \ell_\infty(\boldsymbol{\theta})|. \qquad \text{(A2)}$$

The first term converges by the e.a.s. convergence of $\tilde{f}_t(y^{1:t-1}, \boldsymbol{\theta}, \bar{f}_1)$ to $\tilde{f}_t(y^{t-1}, \boldsymbol{\theta})$ and a continuous mapping argument. The second term converges by the ULLN in Ranga Rao (1962).

For the first term in (A2), we show that $\sup_{\boldsymbol{\theta}\in\Theta} |\tilde{\ell}_t(\boldsymbol{\theta}, \bar{f}_1) - \tilde{\ell}_t(\boldsymbol{\theta})| \xrightarrow{a.s.} 0$ as $t \to \infty$. The expression for the likelihood in the technical appendix of Blasques et al. (2014b) and the maintained differentiability conditions ensure that $\tilde{\ell}_t(\cdot, \bar{f}_1) = \ell(\tilde{f}_t(y^{1:t-1}, \cdot, \bar{f}_1), y_t, \cdot)$ is continuous in $(\tilde{f}_t(y^{1:t-1}, \cdot, \bar{f}_1), y_t)$.

Since the conditions of Corollary 2 hold, and these imply the conditions of Corollary 1, we know that there exists a unique SE sequence $\{\tilde{f}_t(y^{1:t-1}, \boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ such that

$$|\tilde{f}_t(y^{1:t-1}, \boldsymbol{\theta}, \bar{f}_1) - \tilde{f}_t(y^{t-1}, \boldsymbol{\theta})| \xrightarrow{a.s.} 0\ \forall\, \bar{f}_1 \in \mathcal{F}$$

for every $\boldsymbol{\theta} \in \Theta$. In fact, there exists a unique SE sequence $\{f_t(y^{1:t-1}, \cdot)\}_{t\in\mathbb{Z}}$ such that

$$\sup_{\boldsymbol{\theta}\in\Theta} |f_t(y^{1:t-1}, \boldsymbol{\theta}, \bar{f}_1) - f_t(y^{t-1}, \boldsymbol{\theta})| \xrightarrow{a.s.} 0 \quad \forall \bar{f}_1 \in \mathcal{F}$$

and $\sup_t \mathbb{E} \sup_{\boldsymbol{\theta}\in\Theta} |f_t(y^{1:t-1}, \boldsymbol{\theta}, \bar{f}_1)|^{n_f} < \infty$ 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 (A2) 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 (A2), we apply the ergodic theorem for separable Banach spaces of Ranga 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 $\sup_{\boldsymbol{\theta}\in\Theta} |\ell_T(\boldsymbol{\theta}) - \ell_\infty(\boldsymbol{\theta})| \xrightarrow{a.s.} 0$ where $\ell_\infty(\boldsymbol{\theta}) = \mathbb{E}\tilde{\ell}_t(\boldsymbol{\theta})\ \forall\, \boldsymbol{\theta} \in \Theta$. The ULLN $\sup_{\boldsymbol{\theta}\in\Theta} |\ell_T(\boldsymbol{\theta}) - \mathbb{E}\tilde{\ell}_t(\boldsymbol{\theta})| \xrightarrow{a.s.} 0$ 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 $\sup_{\boldsymbol{\theta}\in\Theta} \mathbb{E}|f_t(y^{t-1}, \boldsymbol{\theta})|^{n_f} < \infty\ \forall\, \boldsymbol{\theta} \in \Theta$, $\mathbb{E}|y_t|^{n_y} < \infty$ which holds again by Corollary 2, and the fact that Assumption 4 implies $\ell \in \mathbb{M}(\boldsymbol{n}, n_\ell)$ with $\boldsymbol{n} = (n_f, n_y)$ and $n_\ell \geq 1$; see the technical appendix of Blasques et al. (2014b).

‡Blasques and Lucas thank the Dutch National Science Foundation (NWO; grant VICI453-09-005) for financial support. Koopman acknowledges support from CREATES, Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation. Email correspondence: f.blasques@vu.nl, s.j.koopman@vu.nl, a.lucas@vu.nl.

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. $\square$

Proof of Theorem 2. In this proof we denote the ith order derivative of the filter $\{\tilde{f}_t(\boldsymbol{\theta}, \bar{f}_1)\}_{t\in\mathbb{N}}$ by $\boldsymbol{f}_t^{(i)}(y^{1:t-1}, \boldsymbol{\theta}, \bar{\boldsymbol{f}}_1^{0:i})$, which takes values in $\mathcal{F}^{(i)}$, with $\bar{\boldsymbol{f}}_1^{0:i} \in \mathcal{F}^{(0:i)} = \mathcal{F} \times ... \times \mathcal{F}^{(i)}$ being the fixed initial condition for the first ith order derivatives; see the Supplementary Appendix of Blasques et al. (2014) for further details.

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 $\ell_T(\boldsymbol{\theta}, \bar{f}_1)$ in $\boldsymbol{\theta} \in \Theta$; (iii) the asymptotic normality of the score

$$\sqrt{T}\ell_T'\big(\boldsymbol{\theta}_0, \boldsymbol{f}_1^{(0:1)}\big) \xrightarrow{d} \mathrm{N}(0, \mathcal{J}(\boldsymbol{\theta}_0)), \qquad \mathcal{J}(\boldsymbol{\theta}_0) = \mathbb{E}\big(\tilde{\ell}_t'(\boldsymbol{\theta}_0)\tilde{\ell}_t'(\boldsymbol{\theta}_0)^\top\big); \tag{A3}$$

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

$$\sup_{\boldsymbol{\theta}\in\Theta} \big\|\ell_T''(\boldsymbol{\theta}, \boldsymbol{f}_1^{(0:2)}) - \ell_\infty''(\boldsymbol{\theta})\big\| \xrightarrow{a.s.} 0; \tag{A4}$$

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 technical appendix.

Step 2, CLT: The asymptotic normality of the score in (A6) follows by Theorem 18.10[iv] in Van der Vaart (2000) by showing that,

$$\|\ell_T'\big(\boldsymbol{\theta}_0, \boldsymbol{f}_1^{(0:1)}\big) - \ell_T'\big(\boldsymbol{\theta}_0\big)\| \xrightarrow{e.a.s.} 0 \ \text{ as } \ T \to \infty. \tag{A5}$$

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

$$\sqrt{T}\ell_T'\big(\boldsymbol{\theta}_0\big) \xrightarrow{d} \mathrm{N}(0, \mathcal{J}(\boldsymbol{\theta}_0)\big) \ \text{ as } \ T \to \infty, \tag{A6}$$

where $\mathcal{J}(\boldsymbol{\theta}_0) = \mathbb{E}(\tilde{\ell}_t'\big(\boldsymbol{\theta}_0\big)\tilde{\ell}_t'\big(\boldsymbol{\theta}_0\big)^\top) < \infty$ follows from having $n_{\ell'} \geq 2$; see the expressions for the likelihood in Section B.1 of the technical appendix of Blasques et al. (2014b).

To establish the e.a.s. convergence in (A5), we use the e.a.s. convergence

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

The convergence of the filter directly implied by the fact that the conditions of Corollary 1 hold as they are implied by those of Corollary 2. The convergence of its derivatives holds by Proposition 2 in Blasques et al. (2014) which shows that the derivative processes converge under the same conditions as the filter itself and the maintained moment preservation conditions. From the differentiability of

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

in $\tilde{\boldsymbol{f}}_t^{(0:1)}(y^{1:t-1}, \boldsymbol{\theta}, \boldsymbol{f}_1^{(0:1)})$ and the convexity of $\mathcal{F}$, there exists a set Θ containing $\boldsymbol{\theta}_0$ where we can apply the mean-value theorem to obtain

$$\|\ell_T'\big(\boldsymbol{\theta}_0, \bar{\boldsymbol{f}}_1^{(0:1)}\big) - \ell_T'\big(\boldsymbol{\theta}_0\big)\| \leq \sum_{j=1}^{4+d_\lambda} \Big|\frac{\partial \ell'(y^{1:t}, \hat{\boldsymbol{f}}_t^{(0:1)})}{\partial f_j}\Big|\big|\tilde{\boldsymbol{f}}_{j,t}^{(0:1)}(y^{1:t-1}, \boldsymbol{\theta}_0, \bar{\boldsymbol{f}}_1^{(0:1)}) - \tilde{\boldsymbol{f}}_{j,t}^{(0:1)}(y^{1:t-1}, \boldsymbol{\theta}_0)\big|, \tag{A7}$$

where $\tilde{\boldsymbol{f}}_{j,t}^{(0:1)}$ denotes the j-th element of $\tilde{\boldsymbol{f}}_t^{(0:1)}$, and $\hat{\boldsymbol{f}}^{(0:1)}$ is a point between $\tilde{\boldsymbol{f}}_{j,t}^{(0:1)}(y^{1:t-1}, \boldsymbol{\theta}_0, \bar{\boldsymbol{f}}_1^{(0:1)})$ and $\tilde{\boldsymbol{f}}_{j,t}^{(0:1)}$. By Proposition 2 in Blasques et al. (2014) it follows that under the maintained conditions there exists a set Θ containing $\boldsymbol{\theta}_0$ such that

$$\mathbb{E}\Big|\frac{\partial \ell'(y^{1:t}, \hat{\boldsymbol{f}}_t^{(0:1)})}{\partial f_j}\Big| < \infty$$

and, as a result, $\big|\partial \ell'(y^{1:t}, \hat{\boldsymbol{f}}_t^{(0:1)})/\partial f\big| = O_p(1)$. The strong convergence in (A7) is now ensured by

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

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

$$\sup_{\boldsymbol{\theta}\in\Theta} \|\ell_T''(\boldsymbol{\theta}, \bar{f}_1) - \ell_\infty''(\boldsymbol{\theta})\| \leq \sup_{\boldsymbol{\theta}\in\Theta} \|\ell_T''(\boldsymbol{\theta}, \bar{f}_1) - \ell_T''(\boldsymbol{\theta})\| + \sup_{\boldsymbol{\theta}\in\Theta} \|\ell_T''(\boldsymbol{\theta}) - \ell_\infty''(\boldsymbol{\theta})\|. \tag{A8}$$

To prove that the first term vanishes a.s., we show that $\sup_{\boldsymbol{\theta}\in\Theta} \|\tilde{\ell}_t''(\boldsymbol{\theta}, \bar{f}_1) - \tilde{\ell}_t''(\boldsymbol{\theta})\| \stackrel{a.s.}{\to} 0$ as $t \to \infty$. The differentiability of $\tilde{g}$, $\tilde{g}'$, $\tilde{p}$, and S ensure that

$$\tilde{\ell}_t''(\cdot, \bar{f}_1) = \ell''(y_t, \tilde{\boldsymbol{f}}_t^{(0:2)}(y^{1:t-1}, \cdot, \bar{\boldsymbol{f}}_{0:2}), \cdot)$$

is continuous in $(y_t, \tilde{\boldsymbol{f}}_t^{(0:2)}(y^{1:t-1}, \cdot, \bar{\boldsymbol{f}}_{0:2}))$. Furthermore, the continuous differentiability of these maps ensures also that there exists a set Θ containing $\boldsymbol{\theta}_0$ and a unique SE sequence $\{\tilde{\boldsymbol{f}}_t^{(0:2)}(y^{t-1}, \cdot)\}_{t\in\mathbb{Z}}$ with elements taking values in $\mathbb{C}(\Theta \times \mathcal{F}^{(0:i)})$ such that

$$\sup_{\boldsymbol{\theta}\in\Theta} \big\|(y_t, \tilde{\boldsymbol{f}}_t^{(0:2)}(y^{1:t-1}, \boldsymbol{\theta}, \bar{\boldsymbol{f}}_{0:2})) - (y_t, \tilde{\boldsymbol{f}}_t^{(0:2)}(y^{t-1}, \boldsymbol{\theta})\big\| \stackrel{a.s.}{\to} 0$$

and satisfying, for for $n_f \geq 1$,

$$\sup_t \mathbb{E} \sup_{\boldsymbol{\theta}\in\Theta} \|\tilde{\boldsymbol{f}}_t^{(0:2)}(y^{1:t-1}, \boldsymbol{\theta}, \bar{\boldsymbol{f}}_{0:2})\|^{n_f} < \infty$$

and also $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} \|\tilde{\boldsymbol{f}}_t^{(0:2)}(y^{t-1}, \boldsymbol{\theta})\|^{n_f} < \infty$. The first term in (A8) 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 (A8) 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 under the maintained assumptions on some Θ containing $\boldsymbol{\theta}_0$. The moment bound $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} \|\tilde{\ell}_t''(\boldsymbol{\theta})\| < \infty$ follows from the fact that under the present conditions, Corollary 2 holds, and hence, by Proposition 2 in Blasques et al. (2014), the filtered process and its derivatives have n_f, $n_f^{(1)}$ and $n_f^{(2)}$ moments, respectively, at $\boldsymbol{\theta}_0$. As a result, $n_{\ell''} \geq 1$ ensures $\mathbb{E}\|\tilde{\ell}_t''(\boldsymbol{\theta}_0)\| < \infty$, and by continuity of ℓ'', $\exists$ some Θ containing $\boldsymbol{\theta}_0$ such that $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta} \|\tilde{\ell}_t(\boldsymbol{\theta})\| < \infty$.

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. $\square$