Maximum Likelihood Estimation for Score-Driven Time Series Models *

Francisco Blasques[(a)] Janneke van Brummelen[(a)]
Siem Jan Koopman[(a,b)†] André Lucas[(a)]

[(a)] Vrije Universiteit Amsterdam and Tinbergen Institute
[(b)] CREATES, Aarhus University

September 2020

Abstract

We establish strong consistency and asymptotic normality of the maximum likelihood estimator for stochastic time-varying parameter models driven by the score of the predictive conditional likelihood function. For this purpose, we formulate primitive conditions for global identification, invertibility, strong consistency, and asymptotic normality both under correct specification and misspecification of the model. A detailed illustration is provided for a conditional volatility model with disturbances from the Student's t distribution.

Key words: time-varying parameters, Markov processes, stationarity, invertibility, consistency, asymptotic normality.

JEL classification: C13, C22.

---

*We thank Peter Boswijk, Christian Francq, Andrew Harvey, Anders Rahbek, the editor, associate editor, and two anonymous referees, as well as participants of the "Workshop on Dynamic Models driven by the Score of Predictive Likelihoods", Amsterdam; the "7th International Conference on Computational and Financial Econometrics", London; the "Workshop on Dynamic Models driven by the Score of Predictive Likelihoods", Tenerife; the IAAE London conference; and seminar participants at Vrije Universiteit Amsterdam, University of Cologne, CREST Paris, Maastricht University, for helpful comments and discussions. Blasques and Lucas thank the Dutch National Science Foundation (NWO; grant VICI453-09-005) for financial support. Blasques thanks the Dutch Science Foundation (NWO; grant VI.Vidi.195.099) for financial support. Koopman acknowledges support from CREATES, Aarhus University, Denmark, funded by the Danish National Research Foundation, (DNRF78). An overview of contributions on score-driven models is provided at `http://gasmodel.com`.

†Corresponding author: S.J. Koopman, Vrije Universiteit Amsterdam, School of Business and Economics, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands. Phone: +31205986019, Fax: +31205986020, Email: `s.j.koopman@vu.nl`.

1 Introduction

We study the asymptotic properties of the maximum likelihood estimator for score-driven time series models as introduced by Creal et al. (2011, 2013) and Harvey (2013). We specify the score-driven model as

$$\begin{aligned} y_t &= g(f_t, u_t), \qquad u_t \sim p_u(u_t; \lambda), \qquad f_{t+1} = \omega + \alpha\, s_t + \beta f_t, \\ s_t &= S_t \cdot \nabla_t, \qquad \nabla_t = \partial \log p_y(y_t | f_t; \lambda)/\partial f_t, \end{aligned} \tag{1.1}$$

where y_t denotes the observed data, $g(\cdot, \cdot)$ is a link function that is strictly increasing in its second argument, f_t is a stochastic time-varying parameter that indexes the predictive conditional density p_y of the data y_t, u_t is an independent and identically distributed (i.i.d.) innovation with density p_u, λ is the static parameter vector that indexes p_u, ω, α and β are fixed unknown parameters, s_t is the scaled score function using scaling function $S_t := S(f_t; \lambda)$, and ∇_t is the score (i.e. the derivative of the log) of the predictive conditional density $p_y(y_t|f_t; \lambda)$ with respect to f_t. The conditional (on f_t) density p_y is implied by the innovation density p_u and the link function g. We gather the static parameters in a parameter vector $\boldsymbol{\theta}^\top = (\omega, \alpha, \beta, \lambda^\top)$, where $^\top$ denotes the transpose of a vector or matrix. We estimate $\boldsymbol{\theta}$ by the method of maximum likelihood (ML).

The class of score-driven time series models encompasses many well-known time-varying parameter models from the literature. Additionally, it has given rise to a new strand of literature on successful empirical models in economics and finance. Traditional models contained in the score-driven class include the generalized autoregressive conditional heteroskedasticity (GARCH) model of Engle (1982) and Bollerslev (1986), the autoregressive conditional duration (ACD) model of Engle and Russell (1998), the multiplicative error model (MEM) of Engle (2002), and many more. Among the wide range of new successful empirical models, we have the dynamic models for location and scale of fat-tailed data (Harvey and Luati, 2014), mixed measurement dynamic factor structures (Creal et al., 2014), dynamic models for multivariate count data (Koopman et al., 2018; Babii et al., 2019), dynamic spatial processes (Blasques et al., 2016; Catania and Billé, 2017), dynamic tail indices (Massacci, 2016), and dynamic copulas, both with short-memory dynamics (Creal et al., 2011; Lucas et al., 2014, 2017), long-memory dynamics (Janus et al., 2014), factor structures (Oh and Patton, 2018), and with realized measures as inputs (De Lira Salvatierra and Patton, 2015; Opschoor et al., 2018).

Despite this range of new empirical models with score-driven dynamics, few general theoretical results are available for the asymptotic properties of maximum likelihood estimators in such models. The main complication lies in the nonlinearity of the updating equation

(1.1) in score-driven models. In this paper, we aim to fill this gap by deriving new asymptotic results for the maximum likelihood estimator that are applicable to a wide class of score-driven models.

A distinguishing feature of score-driven time series models is the use of the scaled score s_t in the transition equation for f_{t+1} in (1.1). This makes the model observation-driven in the classification of Cox (1981). Therefore, maximum likelihood estimation of static parameters can be achieved via a prediction error decomposition. In particular, we can express the likelihood function in closed-form, which significantly reduces the computational burden. Blasques et al. (2015) show that score-driven models have unique optimality properties in terms of approximating the unknown sequence of conditional densities $p_y(y_t|f_t;\lambda)$, even when the model is misspecified. Relatedly, Koopman et al. (2016) show that score-driven time-varying parameter models produce similar forecasting precision as parameter-driven state-space models, even if the latter constitute the true DGP.

Our asymptotic results for the maximum likelihood estimator have a number of distinctive features compared to earlier theoretical contributions on observation-driven and in particular score-driven time series models. First, the asymptotic properties that we derive for the maximum likelihood estimator (MLE) are global. For example, we provide a global identification result for score-driven models in terms of low-level conditions. This new result differs from the existing literature that typically relies on high-level assumptions and only ensures local identification by imposing invertibility conditions on the information matrix at the true parameter value; see, for example, Straumann and Mikosch (2006), Meitz and Saikkonen (2011) and Harvey (2013). Second, we formulate primitive *low-level* conditions in terms of the basic structure of the model. For instance, we obtain the required moments of the likelihood function directly from assumptions concerning the properties of the basic building blocks of the model in (1.1), such as the shape of the density function p_y. The use of primitive conditions is typically helpful for empirical researchers who want to establish the asymptotic properties of the MLE for their own model. We are able to obtain low-level conditions by adapting Theorem 3.1 in Bougerol (1993). The adapted theorem delivers the strict stationarity and ergodicity of stochastic sequences and also produces bounded moments for the filter. Third, we follow Straumann and Mikosch (2006) in making use of Theorem 3.1 in Bougerol (1993) and the ergodic theorem in Rao (1962) for strictly stationary and ergodic sequences on separable Banach spaces. Based on these results, we establish the invertibility of the score filter and we obtain asymptotic results under weaker differentiability conditions than the existing literature on MLE for score-driven models. Finally, we explore consistency and asymptotic normality results for both well-specified and misspecified models. These results also extend the literature for score-driven models, which thus far focusses only on the correctly specified case. By allowing for model misspecification, we 'align' the asymptotic estimation theory for score-driven models with the existing information-theoretic optimality results of Blasques et al. (2015).

The theory developed here allows us to establish results for a much wider range of score-driven models than studied in current literature, such as models with fat-tailed log-likelihoods and uniformly bounded third order derivatives; see, for example, Harvey (2013), Harvey and Luati (2014), Caivano and Harvey (2014), and Ryoko (2016). In particular, we emphasize that by establishing the invertibility of the score-driven filter, our asymptotic results stand in sharp contrast to existing results on score-driven models that do not ensure invertibility; see also Andres and Harvey (2012) and Harvey and Lange (2015a,b). The importance of filter invertibility for consistency of the MLE has been underlined in Straumann and Mikosch (2006), Wintenberger (2013), and Blasques et al. (2018), among others. Without invertibility, the existing asymptotic results on score-driven models must implicitly assume that the initial value of the true stochastic time-varying parameter, f_1, is random *and* known exactly, while the remaining sequence $\{f_t\}_{t\geq 2}$ is unobserved. This seems highly unrealistic and unsatisfactory.

The lack of theoretical results for the MLE in score-driven models also stands in sharp contrast to the large number of results available for GARCH models. We do not attempt to review that literature here; for good overviews, see for instance Straumann (2005) or Francq and Zakoïan (2010). The main cause for the limited theoretical progress for score-driven models lies in their complex nonlinear dynamic structure compared to common GARCH models. This results in new theoretical challenges and puzzles. The analysis of score-driven models also provides a different perspective from the standard literature: the characteristics of the likelihood function (based on the conditional density p_y) in a score-driven model hinge directly together with the dynamic properties of the stochastic time-varying parameter (via the use of the score $\partial \log p_y/\partial f_t$ in the transition equation (1.1) for f_t). This provides a close link between the two that departs from most of the literature, where the properties of the likelihood function and those of the time-varying parameter can be dealt with separately.

The notation in the remainder of this paper is at times involved. Therefore, we illustrate all steps by a sufficiently tractable leading example, namely the Student's t conditional volatility model introduced in Creal et al. (2011, 2013). By illustrating all details using this example, we keep the exposition focused. The application of the theory is, however, not limited to this particular case. Additional illustrations include various nonlinear and non-Gaussian models that have been referred to in the discussion above. In particular, we work out a second example in detail in Section 5.

The remainder of the paper is organized as follows. In Section 2 the general modeling framework is presented. In Section 3 we obtain stationarity, ergodicity, invertibility, and bounded moments of filtered score-driven sequences using primitive conditions. In Section 4 we prove our results on global identification, consistency, and asymptotic normality of the MLE. Section 5 provides a detailed application of the results to a fat-tailed score-driven time-varying location model. Concluding remarks can be found in Section 6. The proofs of the main theorems are collected in the Appendix. More technical material is relegated to the Technical Appendix.

2 The General Framework

We develop our asymptotic framework for the score-driven time-varying parameter model in (1.1) in terms of its two main building blocks: the innovation density p_u and the link function g. The conditional data density p_y is implied by p_u and g as follows

$$p_y\big(y_t \,\big|\, f_t\,;\,\lambda\big) = p_u\big(\bar{g}(f_t, y_t)\,;\,\lambda\big)\cdot \bar{g}'(f_t, y_t), \tag{2.1}$$

where all variables are introduced below (1.1), and where

$$\bar{g}_t := \bar{g}(f_t, y_t) := g^{-1}(f_t, y_t), \qquad \bar{g}_t' := \bar{g}'(f_t, y_t) := \partial \bar{g}(f_t, y)/\partial y|_{y=y_t},$$

are the inverse of $g(f_t, u_t)$ with respect to its second argument, u_t, and the Jacobian of the transformation, respectively. We assume $y_t \in \mathcal{Y} \subseteq \mathbb{R}$ and $g : \mathcal{F} \times \mathcal{U} \to \mathcal{Y}$, where $\mathcal{Y}$, $\mathcal{U}$, and $\mathcal{F}$ are the convex domains of y_t, u_t, and f_t, respectively. For ease of exposition, we set the dimension of the parameter vector λ to one. All results can be generalized to the high-dimensional case straightforwardly, because none of the arguments used in the proofs rely on λ being a scalar.

We denote the initialized stochastic time-varying parameter, also called the filtered parameter, by $\hat{f}_t(\boldsymbol{\theta}, \hat{f}_1)$, as it depends on the static parameter vector $\boldsymbol{\theta} = (\omega, \alpha, \beta, \lambda) \in \Theta \subseteq \mathbb{R}^4$ and the non-random initialization $\hat{f}_1 \in \mathcal{F}$ for $t = 1$. For notational simplicity, we suppress the dependence of $\hat{f}_t(\boldsymbol{\theta}, \hat{f}_1)$ on its arguments whenever possible and write $\hat{f}_t$ instead. The stationary limit of $\hat{f}_t$, which does not depend on the initialization $\hat{f}_1$, is denoted by $f_t := f_t(\boldsymbol{\theta})$, where again the argument $\boldsymbol{\theta}$ is usually suppressed.

In case $\boldsymbol{\theta}_0$ is the true static parameter, then $f_t(\boldsymbol{\theta}_0)$ is the true stochastic time-varying parameter driving the model. We assume that the true time-varying parameter originates in the infinite past, and hence, has no initialization. A similar approach is found in settings in which the process can be unfolded backwards in time and is shown to converge to a stationary sequence that extends to the infinite past. Typical examples include linear autoregressive models, threshold autoregressive models, GARCH models and autoregressive models with random coefficients. The work of Bougerol (1993) and Straumann and Mikosch (2006) rely on the same assumption.

We adopt a leading example throughout the expositions below to explain the notation and to illustrate how our results can be applied in a concrete setting.

Main example. *Consider the Student's t based time-varying scale model. This model was originally proposed by Creal et al. (2011, 2013) and Harvey (2013) in the context of modeling daily financial returns, and encompasses the celebrated GARCH model of Engle (1982) and Bollerslev (1986). The model is given by*

$$y_t = f_t^{1/2} \cdot u_t, \tag{2.2}$$

where u_t *is an innovation; the model is a special case of (1.1) with* $g(f_t, u_t) = f_t^{1/2} \cdot u_t$*, which is strictly increasing in* u_t *if* $f_t > 0$*, and with Student's t density* p_u*. We also obtain* $\bar{g}_t = g^{-1}(f_t, y_t) = y_t / f_t^{1/2}$ *and* $\bar{g}_t' = f_t^{-1/2}$*. To ensure positivity of the scale* f_t *for all* t*, we impose* $\beta \geq \alpha \geq 0$*,* $\omega > 0$*, and* $\hat{f}_1 > 0$*, where* $\hat{f}_1$ *is the initial condition for* f_t *at time* $t = 1$*. It follows that*

$$p_y(y_t|f_t; \lambda) = \frac{\Gamma\Big((\lambda+1)/2\Big)}{\Gamma(\lambda/2)\sqrt{\pi\,\lambda\, f_t}} \, \Big(1 + \lambda^{-1} y_t^2 / f_t\Big)^{-(\lambda+1)/2}.$$

where λ *is the degrees of freedom parameter. The time-varying parameter* f_t *should not be interpreted as a variance, because we do not impose* $\mathbb{E}[u_t^2] = 1$*. Instead,* f_t *can be viewed as a scaling parameter and we can refer to the model as a conditional scaling model. Creal et al. (2011, 2013), discuss a case where* f_t *can be interpreted as a variance because a scaled Student's t distribution is used for the innovation* u_t*.*

The characteristic feature of score-driven models is their use of the scaled score function as the driving mechanism in transition equation (1.1). Given the decomposition (2.1), we obtain

$$\nabla_t(f_t, y_t; \lambda) = \left[\frac{\partial \bar{p}_t}{\partial f} + \frac{\partial \log \bar{g}_t'}{\partial f}\right]\Bigg|_{f=f_t}, \tag{2.3}$$

with $\bar{p}_t := \bar{p}(f_t, y_t; \lambda) = \log p_u(\bar{g}(f_t, y_t); \lambda)$ and $\bar{g}_t' = \partial \bar{g}(f_t, y)/\partial y|_{y=y_t}$. The scaling function $S : \mathcal{F} \times \Lambda \to \mathcal{F}$ in (1.1) should be positive. Often, it is taken as a power of the conditional inverse Fisher information to account for the curvature of the score at time t; see Creal et al. (2013) for more details.

Main example (continued). *Given the Student's t density p_y, we obtain updating function*

$$\hat{f}_{t+1} = \omega + \alpha \left(w_t\, y_t^2 - \hat{f}_t \right) + \beta \hat{f}_t,$$
$$w_t = (1+\lambda^{-1})/(1+\lambda^{-1}\, y_t^2/\hat{f}_t), \tag{2.4}$$

for nonrandom initial value $\hat{f}_1$, where we used a scaling function $S(\hat{f}_t;\lambda) = 2\hat{f}_t^2$ proportional to the inverse conditional Fisher information. The score-driven scale dynamics in (2.4) have the interesting feature that they downweight large realizations y_t via the weights w_t in (2.4). It gives the score-driven model the desirable robustness feature that is lacking in the GARCH model with Student's t distributed innovations; see Creal et al. (2011) and Harvey and Luati (2014) for more details.

For the limiting case $\lambda \to \infty$, y_t becomes conditionally normally distributed, and we recover a slightly reparameterized version of the standard GARCH model of Engle (1982) and Bollerslev (1986), $\hat{f}_{t+1} = \omega + \alpha y_t^2 + (\beta - \alpha)\hat{f}_t$. For finite λ, however, the recursion in (2.4) is highly nonlinear in both y_t and $\hat{f}_t$.

Section 4 establishes the asymptotic properties of the maximum likelihood estimator (MLE) for the static parameter vector $\boldsymbol{\theta}$. We define the MLE $\hat{\boldsymbol{\theta}}_T(\hat{f}_1)$ for fixed initial condition $\hat{f}_1$ as

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

with the average log-likelihood function ℓ_T given in closed form as

$$\ell_T(\boldsymbol{\theta}, \hat{f}_1) = \frac{1}{T}\sum_{t=1}^{T}\left(\log p_u(\bar{g}(\hat{f}_t, y_t);\lambda) + \log \frac{\partial \bar{g}(\hat{f}_t, y_t)}{\partial y} \right) = \frac{1}{T}\sum_{t=1}^{T}\left(\bar{p}_t + \log \bar{g}_t' \right). \tag{2.5}$$

The availability of a closed-form expression for the likelihood function is one of the computational advantages of observation-driven time-varying parameter models. It has for instance led to the widespread application of GARCH models in applied empirical work. As is clear from equation (2.5), score-driven models benefit from the same computational advantages.

3 Stochastic Properties of Score-Driven Filters

Before we develop the asymptotic properties of the MLE, we first establish the stationarity, ergodicity, and invertibility properties and the existence of moments of the stochastic time-varying parameter process $\{f_t\}$. We do so using primitive conditions. The likelihood function (2.5) is formulated in terms of the data and in terms of the filtered time-varying parameter $\hat{f}_t$ as defined by the recursion in (1.1). In order for the likelihood function to be well-behaved and for an appropriate law of large numbers (LLN) and central limit theorem (CLT) to apply, the *filtered* sequence $\{\hat{f}_t\}$ as well as the sequences of its first and second order derivatives need to be sufficiently well-behaved for a given data sequence $\{y_t\}$. Naturally, the filtered $\{\hat{f}_t\}$ sequence for given data $\{y_t\}$ needs to be carefully distinguished from its model-implied counterpart, which takes the innovations $\{u_t\}$ rather than the data $\{y_t\}$ as given. We will therefore denote this later sequence by $\{f_t^u\}$. In this section we investigate the properties of both the filtered and model-implied sequences. The results below are used in Section 4 to establish the asymptotic properties of the MLE for $\boldsymbol{\theta}$.

We first introduce some additional notation. For a scalar random variable x, define $\|x\|_n := (\mathbb{E}|x|^n)^{1/n}$ for $n > 0$. If the random variable $x(\boldsymbol{\theta})$ depends on a parameter $\boldsymbol{\theta} \in \Theta$, define $\|x(\cdot)\|_n^\Theta := (\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|x(\boldsymbol{\theta})|^n)^{1/n}$. We say that the sequence $\{x_t\}$ converges exponentially fast almost surely (e.a.s.) to the sequence $\{x_t'\}$ if $c^t\,\|x_t - x_t'\| \overset{a.s.}{\to} 0$ for some $c > 1$; see Straumann and Mikosch (2006).

Propositions 3.1 and 3.3 below are written specifically for the score-driven recursion in (1.1). The propositions can, however, be extended to more general forms which can be found in Technical Appendix B. First, we consider the score-driven model defined in terms of the innovations u_t rather than in terms of the observations y_t. This enables us to establish explicit results for the score-driven model as a potential data generating process and to derive properties for the MLE under the assumption of a correctly specified model. Define $s_{u,t} := s(\hat{f}_t^u, g(\hat{f}_t^u, u_t); \lambda)$ where $s(f_t, y_t; \lambda) = S(f_t; \lambda) \cdot \nabla_t(f_t, y_t; \lambda)$ and let $\{\hat{f}_t^u\}_{t\in\mathbb{N}}$ be generated by

$$\hat{f}_{t+1}^u = \omega + \alpha\, s_{u,t} + \beta\, \hat{f}_t^u, \tag{3.1}$$

for $t > 1$ and an initial non-random value $\hat{f}_1^u \in \mathbb{R}$.

Main example (continued). *The recursion in* (2.4) *is defined in terms of* y_t *and* $\hat{f}_t$. *If we define the recursion in terms of* u_t *and* $\hat{f}_t^u$ *instead as required by equation* (3.1)*, we obtain*

$$\hat{f}_{t+1}^u = \omega + \left(\beta + \alpha\left(\frac{(1+\lambda^{-1})u_t^2}{1+\lambda^{-1}u_t^2} - 1\right)\right) \cdot \hat{f}_t^u, \tag{3.2}$$

such that $s_{u,t} = ((1+\lambda^{-1})u_t^2/(1+\lambda^{-1}u_t^2) - 1) \cdot \hat{f}_t^u$. *So whereas the recursion in* (2.2) *is highly nonlinear in* $\hat{f}_t$ *given* y_t*, the recursion in* (3.2) *is linear in* $\hat{f}_t^u$ *for given* u_t. *The data generating process* (3.2) *allows large values of* u_t *to have a small impact on the stochastic time-varying parameter, relative to a GARCH model with Student's t innovations. The robustness feature of* (3.2) *is well suited for a model with heavy-tailed innovations.*

We next formulate a result for the stationarity and existence of moments of $\{f_t^u\}_{t\in\mathbb{Z}}$, the limit process of $\{\hat{f}_t^u\}_{t\in\mathbb{N}}$ as given by (3.1). This generalizes the results of Blasques et al. (2018) which establishes only stationarity, but not bounded moments, and the results of

Blasques et al. (2014) which hold only for the special case of dynamic correlation models. We assume that the scaled score s_u is continuously differentiable in f_t^u and continuous in u_t and λ. Define

$$\rho_t^k(\boldsymbol{\theta}) := \sup_{f^{u*}\in\mathcal{F}} \left|\, \beta + \alpha\, \partial s_{u,t}/\partial f^u|_{f^u=f^{u*}}\, \right|^k. \tag{3.3}$$

We then have the following proposition.

Proposition 3.1. *For every $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^4$ let $\{u_t\}_{t\in\mathbb{Z}}$ be an i.i.d. sequence and assume $\exists \hat{f}_1^u \in \mathcal{F}$ such that*

(i) $\mathbb{E}\log^+ |s_u(\hat{f}_1^u, u_1; \lambda)| < \infty$*;*

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

Then $\{\hat{f}_t^u\}_{t\in\mathbb{N}}$ converges exponentially fast almost surely (e.a.s.) to the unique stationary and ergodic (SE) sequence $\{f_t^u\}_{t\in\mathbb{Z}}$ for every $\boldsymbol{\theta} \in \Theta$ as $t \to \infty$.

If furthermore for every $\boldsymbol{\theta} \in \Theta$ there exists some $n_{f^u} > 0$ such that

(iii) $\|s_u(\hat{f}_1^u, u_1; \lambda)\|_{n_{f^u}} < \infty$*;*

(iv) $\mathbb{E}\rho_1^{n_{f^u}}(\boldsymbol{\theta}) < 1$*;*

then $\mathbb{E}|f_t^u|^{n_{f^u}} < \infty$.

Proposition 3.1 not only establishes stationarity and ergodicity (SE) of f_t^u, it also establishes existence of unconditional moments. Furthermore, conditions *(i)* and *(ii)* in Proposition 3.1 also provide an almost sure representation of f_t^u in terms of $\{u_t\}_{t=-\infty}^{t-1}$. We refer to the Technical Appendix for further details.

Remark 3.2. Proposition 3.1 also holds if the supremum in (3.3) is defined over a larger convex set $\mathcal{F}^* \supseteq \mathcal{F}$. The same holds for Proposition 3.3 later on. This can for instance be used if the original space $\mathcal{F}$ is non-convex.

Main example (continued). *In our main example, the recursion (3.1) is always linear in $\hat{f}_t^u$; see equation (3.2). Conditions (i) and (iii) are satisfied for $0 < \lambda < \infty$ because $(1+\lambda^{-1})u_t^2/(1+\lambda^{-1}u_t^2)$ is uniformly bounded in u_t by the constant $\lambda + 1 < \infty$. Conditions (ii) and (iv) are satisfied if the factor in front of $\hat{f}_t^u$ in (3.2) has a log-moment or an n_{f^u} moment, respectively. For example, for $n_{f^u} = 1$ condition (iv) collapses to $0 < \beta < 1$.*

Proposition 3.1 will prove convenient in case the model is correctly specified as it describes the properties of the score-driven model as a data generating process as well as the properties of the score filter at the true parameter $\boldsymbol{\theta}_0 \in \Theta$. We namely have that $f_t^u(\boldsymbol{\theta}_0) = f_t(\boldsymbol{\theta}_0)$.

Irrespective of whether we have a correct or an incorrect specification of the model, to derive the MLE properties we must always analyze the stochastic behavior of the filtered time-varying parameter over different $\boldsymbol{\theta} \in \Theta$. Proposition 3.3 presented below is key in establishing the invertibility, moment bounds and e.a.s. convergence uniformly over the parameter space Θ of the score-driven filtered sequence $\{\hat{f}_t\}$, formulated in terms of the data $\{y_t\}$ rather than in terms of the innovations $\{u_t\}$ as in equation (3.1). We assume that s is differentiable in f_t and continuous in y_t and λ. To state our subsequent proposition concisely, we define the supremum

$$\bar{\rho}_t(\boldsymbol{\theta}) = \sup_{f^* \in \mathcal{F}} \Big| \, \beta + \alpha \, \partial s(f, y_t; \lambda)/\partial f|_{f=f^*} \, \Big|. \tag{3.4}$$

Proposition 3.3. *Let $\Theta \subset \mathbb{R}^4$ be compact, and let $\{y_t\}_{t\in\mathbb{Z}}$ be an SE sequence. Assume $\exists \; \hat{f}_1 \in \mathcal{F}$ such that*

(i) $\mathbb{E} \log^+ \sup_{\lambda \in \Lambda} |s(\hat{f}_1, y_t; \lambda)| < \infty$;

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

Then the sequence $\{\hat{f}_t\}_{t\in\mathbb{N}}$ converges e.a.s. to a unique limit SE sequence $\{f_t\}_{t\in\mathbb{Z}}$ as $t \to \infty$, uniformly on Θ.

If furthermore $\exists \; n_f > 0$ such that

(iii) $\|s(\hat{f}_1, y_t; \cdot)\|_{n_f}^{\Lambda} < \infty$;

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

then $\|f_t\|_{n_f}^{\Theta} < \infty$.

The conditions of Proposition 3.3 are easily satisfied by many models including score-driven volatility models and time-varying location models, both with different innovation distributions. Specific examples are the one in our main example, the logistic time-varying mean models and the log volatility models with Student's t distributed innovations.

Main example (continued). *Consider the time-varying scale model in equation (2.2) with $0 < \underline{\lambda} \leq \lambda \leq \bar{\lambda} < \infty$. From the uniform boundedness of the score in y_t for given $\hat{f}_1$, we obtain that conditions (i) and (iii) of Proposition 3.3 are trivially satisfied. Furthermore, we have*

$$\dot{s}_{y,t}(f^*; \lambda) \;=\; \frac{\partial s(f, y_t; \lambda)}{\partial f}\bigg|_{f=f^*} \;=\; \frac{(1+\lambda^{-1}) y_t^4 / (\lambda f^{*2})}{\left(1 + \lambda^{-1} y_t^2 / f^*\right)^2} - 1. \tag{3.5}$$

For fixed λ and y_t, $\dot{s}_{y,t}(f^*;\lambda)$ is decreasing in f^*. It attains the value λ as $f^* \to 0$ and the value -1 as $f^* \to \infty$. Given the parameter restriction $\beta \geq \alpha \geq 0$, it follows that $\beta + \alpha\dot{s}_{y,t}(f^*;\lambda) \geq 0$ for every $f^* \in \mathcal{F}$, implying that its absolute value attains its maximum as $f^* \to 0$. Thus, $\sup_{\boldsymbol{\theta}\in\Theta} \bar{\rho}_1(\boldsymbol{\theta}) \leq \sup_{(f^*,y,\boldsymbol{\theta})\in\mathcal{F}\times\mathcal{Y}\times\Theta} |\beta + \alpha \frac{\partial s(f^*,y;\lambda)}{\partial f}| \leq \sup_{\boldsymbol{\theta}\in\Theta} \beta + \lambda\alpha$ and conditions (ii) and (iv) simplify to

$$\sup_{\boldsymbol{\theta}\in\Theta} \beta + \lambda\alpha < 1. \tag{3.6}$$

If this condition is met, then n_f can be set arbitrarily high.

Propositions 3.1 and 3.3 are similar to the results found in Meitz and Saikkonen (2011). In particular, these results are based on Bougerol (1993, Theorem 3.1) and Straumann and Mikosch (2006, Theorem 2.8). The main differences relate only to the specific contexts under consideration.

Conditions *(iii)* and *(iv)* in Proposition 3.3 imply conditions *(i)* and *(ii)*, respectively. We emphasize that under conditions *(i)* and *(ii)* our score filter is *invertible* since we are able to write f_t as a measurable function of all past observations. Most importantly, the invertibility property ensures that the effect of the initialization $\hat{f}_1$ vanishes as $t \to \infty$, and that the filter converges to a unique limit process independently of $\hat{f}_1$; see, for example, Granger and Andersen (1978), Straumann and Mikosch (2006), Wintenberger (2013) and Blasques et al. (2018). Establishing invertibility is usually one of the main challenges for nonlinear time series models with stochastic time-varying parameters.

In Section 4 we show that the stochastic recurrence approach followed in Propositions 3.1 and 3.3 allows us to obtain consistency and asymptotic normality under weaker differentiability conditions than those typically imposed in the score-driven literature; see also Section 2.3 of Straumann and Mikosch (2006). In particular, instead of relying on the usual pointwise convergence plus the stochastic equicontinuity of Andrews (1992) and Pötscher and Prucha (1994), we can obtain uniform convergence through the application of the ergodic theorem of Rao (1962) for sequences in separable Banach spaces. This constitutes a crucial simplification as working with the third order derivatives of the likelihood of a general score-driven model is typically quite cumbersome. We emphasize that alternative uniform convergence results for proving consistency and asymptotic normality have been used before by, amongst others, Straumann and Mikosch (2006), Hafner and Preminger (2009) and Meitz and Saikkonen (2011).

In the remainder of this section we extend the results of Proposition 3.3 to the derivative processes $\partial f_t/\partial\boldsymbol{\theta}$ and $\partial^2 f_t/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top$. We use stationarity, ergodicity, invertibility and bounded moments of the derivative processes for proving the asymptotic normality of the MLE. To simplify notation, we let $\boldsymbol{f}_t^{(i)} \in \mathcal{F}^{(i)}$ denote a vector containing all the ith order derivatives of f_t with respect to $\boldsymbol{\theta}$, where $\hat{\boldsymbol{f}}_1^{(0:i)} \in \mathcal{F}^{(0:i)}$ contains the fixed initial condition for f_t and its derivatives up to order i. Similarly, $\boldsymbol{f}_t^{(0:i)} \in \mathcal{F}^{(0:i)} = \mathcal{F} \times \ldots \times \mathcal{F}^{(i)}$ denotes a vector containing f_t as well as its derivatives with respect to $\boldsymbol{\theta}$ up to order i.

We introduce more elaborate notation to clarify whether we are working with a perturbed sequence or not. The perturbed sequence $\{\hat{\hat{\boldsymbol{f}}}_t^{(i)}\}_{t\in\mathbb{N}}$, where $\hat{\hat{\boldsymbol{f}}}_t^{(i)} := \hat{\hat{\boldsymbol{f}}}_t^{(i)}(\boldsymbol{\theta}, \hat{\boldsymbol{f}}_1^{(0:i)})$, is initialized at $\hat{\boldsymbol{f}}_1^{(0:i)}$ and depends on the non-stationary initialized sequences $\{\hat{f}_t\}_{t\in\mathbb{N}}$ and $\{\hat{\hat{\boldsymbol{f}}}_t^{(1:i-1)}\}_{t\in\mathbb{N}}$, which are only stationary in the limit. The unperturbed initialized sequence $\{\hat{\boldsymbol{f}}_t^{(i)}\}_{t\in\mathbb{N}}$ with $\hat{\boldsymbol{f}}_t^{(i)} = \hat{\boldsymbol{f}}_t^{(i)}(\boldsymbol{\theta}, \hat{\boldsymbol{f}}_1^{(i)})$ instead depends on the limit SE filter $\{\boldsymbol{f}_t^{(0:i-1)}\}_{t\in\mathbb{Z}}$, and is initialized at some $\hat{\boldsymbol{f}}_1^{(i)}$. Under certain conditions, the sequence $\{\hat{\hat{\boldsymbol{f}}}_t^{(i)}\}_{t\in\mathbb{N}}$ converges to the SE unperturbed limit sequence $\{\boldsymbol{f}_t^{(i)}\}_{t\in\mathbb{Z}}$, where $\boldsymbol{f}_t^{(i)} := \boldsymbol{f}_t^{(i)}(\boldsymbol{\theta})$, which depends on the limit SE filter $\{\boldsymbol{f}_t^{(0:i-1)}\}_{t\in\mathbb{Z}}$. Furthermore, in order to work with primitive conditions we use the notion of moment preserving maps, which we define as follows.

Definition 3.4. (Moment Preserving Maps)
A function $h : \mathbb{R}^q \times \Theta \to \mathbb{R}$ is said to be $\boldsymbol{n}/n$-moment preserving, denoted as $h(\cdot;\boldsymbol{\theta}) \in \mathbb{M}_{\Theta_1,\Theta_2}(\boldsymbol{n}, n)$, if and only if $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta_1} |x_{i,t}(\boldsymbol{\theta})|^{n_i} < \infty$ for $\boldsymbol{n} = (n_1, \ldots, n_q)$ and $i = 1, \ldots, q$ implies $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta_2} |h(\boldsymbol{x}_t(\boldsymbol{\theta}); \boldsymbol{\theta})|^n < \infty$. If Θ_1 or Θ_2 consists of a singleton, we replace Θ_1 or Θ_2 in the notation by its single element, e.g., $\mathbb{M}_{\boldsymbol{\theta}_1,\Theta_2}$ if $\Theta_1 = \{\boldsymbol{\theta}_1\}$.

Moment preservation is a natural requirement in proofs of the asymptotic properties of the MLE, because the likelihood and its derivatives are nonlinear functions of the original data y_t, the time varying parameter f_t, and partial derivatives of the score, such as $\partial s(f_t, y; \lambda)/\partial\lambda$ and $\partial^2 s(f_t, y; \lambda)/\partial f_t \partial\lambda$. Bounding moments of the former can thus be accomplished by bounding moments of the latter plus invoking a moment preservation property. Moment preservation is accomplished, for instance, for polynomial functions $h(x; \boldsymbol{\theta}) = \sum_{j=0}^{J} \theta_j x^j \;\forall\, (x, \boldsymbol{\theta}) \in \mathcal{X} \times \Theta$, $\boldsymbol{\theta} = (\theta_0, \ldots, \theta_J) \in \Theta \subseteq \mathbb{R}^J$. It is then trivial to establish $h \in \mathbb{M}_{\boldsymbol{\theta},\boldsymbol{\theta}}(n, m)$ with $m = n/J \;\forall\, \boldsymbol{\theta} \in \Theta$. If Θ is compact, then also $h \in \mathbb{M}_{\Theta,\Theta}(n, m)$ with $m = n/J$. Similarly, every k-times continuously differentiable function $h(\cdot; \boldsymbol{\theta}) \in \mathbb{C}^k(\mathcal{X}) \;\; \forall\, \boldsymbol{\theta} \in \Theta$, with bounded k-th derivative $\sup_{x\in\mathcal{X}} |h^{(k)}(x; \boldsymbol{\theta})| \leq \bar{h}_k(\boldsymbol{\theta}) < \infty \;\forall\, \boldsymbol{\theta} \in \Theta$, satisfies $h \in \mathbb{M}_{\boldsymbol{\theta},\boldsymbol{\theta}}(n, m)$ with $m = n/k \;\forall\, \boldsymbol{\theta} \in \Theta$. If furthermore $\sup_{\boldsymbol{\theta}\in\Theta} \bar{h}_k(\boldsymbol{\theta}) \leq \bar{\bar{h}} < \infty$, then $h \in \mathbb{M}_{\Theta,\Theta}(n, m)$ with $m = n/k$. The Technical Appendix provides further details and examples of moment preserving maps. We note that $\mathbb{M}_{\Theta',\Theta'}(\boldsymbol{n}, n) \subseteq \mathbb{M}_{\Theta,\Theta}(\boldsymbol{n}, n^*)$ for all $n^* \leq n$, and all $\Theta \subseteq \Theta'$.

Using this notation, we let $s \in \mathbb{M}_{\Theta,\Theta}(\boldsymbol{n}, n_s)$ where $\boldsymbol{n} = (n_f, n_y)$, and hence n_s denotes the number of bounded moments of the scaled score $\sup_{\boldsymbol{\theta}\in\Theta} s(f_t, y; \lambda)$, when f_t and y_t have n_f and n_y moments, respectively, uniformly in $\boldsymbol{\theta}$. Furthermore, as a convention, we let n_s^{λ} and $n_s^{f\lambda}$ denote the number of bounded moments for the partial derivatives $\partial s(f_t, y; \lambda)/\partial\lambda$

and $\partial^2 s(f_t, y; \lambda)/\partial f_t \partial\lambda$, respectively, when their arguments have n_f and n_y moments. Also, for the moments of all functions, the argument f_t is always understood to be the stationarity limit filter which has $n_f > 0$ moments under appropriate conditions stated in Proposition 3.3. We shall make extensive use of analogous definitions for other functions and their corresponding partial derivatives. Finally, $\bar{n}$ denotes moments of functions after taking the supremum over f_t. For example, $\bar{n}_s^{f\lambda}$ denotes the number of moments of the random variable $\sup_f |\partial^2 s(f, y; \lambda)/\partial f \partial\lambda|$, uniformly in $\boldsymbol{\theta} \in \Theta$, or in moment preserving notation

$$\sup_f \left| \frac{\partial^2 s(f, y; \cdot)}{\partial f \partial \lambda} \right| \in \mathbb{M}_{\Theta,\Theta}(\boldsymbol{n}, \bar{n}_s^{f\lambda}),$$

with $\boldsymbol{n} = (n_f, n_y)$. We apply the same notational principle to other functions and derivatives.

Proposition 3.5. *Let the conditions of Proposition 3.3 hold with some $n_f > 0$ and suppose that $s \in \mathbb{C}^{(2,0,2)}(\mathcal{F} \times \mathcal{Y} \times \Lambda)$.*

Let $\min\{n_s, n_s^\lambda, \bar{n}_s^f, \bar{n}_s^{\lambda f}, \bar{n}_s^{ff}\} > 0$. Then $\{\hat{\hat{\boldsymbol{f}}}_t^{(1)}\}_{t\in\mathbb{N}}$ converges e.a.s. to a unique SE sequence $\{\boldsymbol{f}_t^{(1)}\}_{t\in\mathbb{Z}}$, uniformly in Θ, and furthermore, we have $\|\boldsymbol{f}_t^{(1)}\|_{n_{f\boldsymbol{\theta}}}^{\Theta} < \infty$ for any $n_{f\boldsymbol{\theta}}$ satisfying

$$n_{f\boldsymbol{\theta}} \le \min\left\{n_f \ , \ n_s \ , \ n_s^\lambda\right\}.$$

If additionally $\min\{n_s^{\lambda\lambda}, \bar{n}_s^{\lambda\lambda f}, \bar{n}_s^{\lambda ff}, \bar{n}_s^{fff}\} > 0$, then the second derivative $\{\hat{\hat{\boldsymbol{f}}}_t^{(2)}\}_{t\in\mathbb{N}}$ converges e.a.s. to a unique SE sequence $\{\boldsymbol{f}_t^{(2)}\}_{t\in\mathbb{Z}}$, uniformly in Θ. Furthermore, we have $\|\boldsymbol{f}_t^{(2)}\|_{n_{f\boldsymbol{\theta\theta}}}^{\Theta} < \infty$ for any $n_{f\boldsymbol{\theta\theta}}$ satisfying

$$n_{f\boldsymbol{\theta\theta}} \le \min\left\{n_{f\boldsymbol{\theta}} \ , \ n_s^{\lambda\lambda} \ , \ \frac{n_s^f n_{f\boldsymbol{\theta}}}{n_s^f + n_{f\boldsymbol{\theta}}} \ , \ \frac{n_s^{ff} n_{f\boldsymbol{\theta}}}{2n_s^{ff} + n_{f\boldsymbol{\theta}}} \ , \ \frac{n_s^{f\lambda} n_{f\boldsymbol{\theta}}}{n_s^{f\lambda} + n_{f\boldsymbol{\theta}}}\right\}.$$

The expressions for $n_{f\boldsymbol{\theta}}$ and $n_{f\boldsymbol{\theta\theta}}$ may appear complex at first sight. However, they arise naturally from expressions for the derivative of f_t with respect to $\boldsymbol{\theta}$. We next analyze the moment conditions of Proposition 3.5 in the practical setting of our main example.

Main example (continued). *For our main example, we obtain from Proposition 3.3 that the limit filtered process f_t has n_f moments for n_f arbitrarily high, uniformly in $\boldsymbol{\theta}$, as long as contraction condition (3.6) is satisfied. So as long as $\sup_{\boldsymbol{\theta}\in\Theta} \beta + \lambda\alpha < 1$ and $\beta \ge \alpha \ge 0$, we can set n_f arbitrarily high.*

The two remaining conditions required in Proposition 3.5 are $\min\{n_s, n_s^\lambda, \bar{n}_s^f, \bar{n}_s^{\lambda f}, \bar{n}_s^{ff}\} > 0$ and $\min\{n_s^{\lambda\lambda}, \bar{n}_s^{\lambda\lambda f}, \bar{n}_s^{\lambda ff}, \bar{n}_s^{fff}\} > 0$. We note that in our main example s_t is uniformly bounded in y_t for fixed f_t. We therefore easily obtain $n_s = n_f$. The remaining derivatives *are straightforward to check as well and can be found in the Technical Appendix. We obtain* $n_s, n_s^\lambda = n_f$*, and* $\bar{n}_s^f, \bar{n}_s^{\lambda f}, \bar{n}_s^{ff} \to \infty$*. As a result,* $\min\{n_s, n_s^\lambda, \bar{n}_s^f, \bar{n}_s^{\lambda f}, \bar{n}_s^{ff}\} = n_f > 0$*. Similarly,* $\min\{n_s^{\lambda\lambda}, \bar{n}_s^{\lambda\lambda f}, \bar{n}_s^{\lambda ff}, \bar{n}_s^{fff}\} = n_f > 0$*, because* $n_s^{\lambda\lambda} = n_f$ *and* $\bar{n}_s^{\lambda\lambda f}, \bar{n}_s^{\lambda ff}, \bar{n}_s^{fff} \to \infty$*.*

Using these results, we obtain $n_{f_{\boldsymbol{\theta}}} \le \min\{n_f, n_s, n_s^\lambda\} = n_f$ *unconditional moments for the first derivative process, and* $n_{f_{\boldsymbol{\theta}\boldsymbol{\theta}}} < \min\{n_{f_{\boldsymbol{\theta}}}, n_f, n_{f_{\boldsymbol{\theta}}}, \frac{1}{2}n_{f_{\boldsymbol{\theta}}}, n_{f_{\boldsymbol{\theta}}}\} = \frac{1}{2}n_{f_{\boldsymbol{\theta}}} \le \frac{1}{2}n_f$ *for the second derivative process. Here we used the fact that for instance* $n_s^f \ge \bar{n}_s^f \to \infty$*.*

Since n_f *can be set arbitrarily high, we can establish moments up to a large order for both derivative processes of the score-driven scale model.*

We emphasize that the moment conditions stated in Proposition 3.5 are primitive in the sense that they relate directly to the basic building blocks of the score filter: the score function and its derivatives. For the practitioner who wishes to verify moment conditions for any given score model, Technical Appendix G provides a detailed compendium of the moment preserving properties of different classes of functions to simplify the verification of the primitive moment conditions in Proposition 3.5. These include examples of robust volatility filtering and robust trend-extraction models, but also standard regression models with time-varying regression coefficients.

4 Identification, Consistency, Asymptotic Normality

Next we formulate conditions under which the MLE is strongly consistent and asymptotically normal. The low-level conditions that we formulate relate directly to the propositions from Section 3. We obtain asymptotic results for the MLE that hold for possibly misspecified models. These results take the properties of observed data as given. In addition, we also obtain asymptotic properties for the MLE that hold for correctly specified models. The latter results require additional conditions designed to ensure that the score model also behaves well as a data generating process. For correctly specified models, we are also able to prove a new global identification result building on low-level conditions rather than on typical high-level assumptions. We defer a short discussion on the usefulness of asymptotic results under strong forms of misspecification until directly after Theorem 4.6 below

We start with two rather standard assumptions.

Assumption 4.1. $(\Theta, \mathfrak{B}(\Theta))$ *is a measurable space and* Θ *is compact.*

Assumption 4.2. $\bar{g} \in \mathbb{C}^{(4,1)}(\mathcal{F} \times \mathcal{Y})$*,* $\bar{g}' \in \mathbb{C}^{(4,0)}(\mathcal{F} \times \mathcal{Y})$*,* $\bar{p} \in \mathbb{C}^{(4,2)}(\tilde{\mathcal{U}} \times \Lambda)$*, and* $S \in \mathbb{C}^{(3,2)}(\mathcal{F} \times \Lambda)$*, where* $\tilde{\mathcal{U}} := \bar{g}(\mathcal{Y}, \mathcal{F})$[1]*.*

[1] The notation used here is ambiguous about the existence of cross-derivatives. Therefore, we impose that

The conditions in Assumption 4.2 are necessary for asymptotic normality of the MLE. Notice that less restrictive assumptions would suffice for existence and consistency of the MLE. For example, for existence continuity in f_t and measurability in y_t would be sufficient.

Let Ξ be the event space of the underlying complete probability space. The next theorem establishes the existence of the MLE.

Theorem 4.3. (Existence) *Let Assumptions 4.1 and 4.2 hold. Then there exists a.s. a measurable map $\hat{\boldsymbol{\theta}}_T : \Xi \to \Theta$ satisfying $\hat{\boldsymbol{\theta}}_T \in \arg\max_{\boldsymbol{\theta}\in\Theta} \ell_T(\boldsymbol{\theta}, \hat{f}_1)$, for all $T \in \mathbb{N}$ and every initialization $\hat{f}_1 \in \mathcal{F}$, where ℓ_T is the average log-likelihood function defined in (2.5).*

Using our notation for moment-preserving maps, let $\log \bar{g}' \in \mathbb{M}_{\Theta,\Theta}(\boldsymbol{n}, n_{\log \bar{g}'})$ and $\bar{p} \in \mathbb{M}_{\Theta,\Theta}(\boldsymbol{n}, n_{\bar{p}})$ as defined below (2.1) and (2.3), respectively, where $\boldsymbol{n} := (n_f, n_y)$. Similarly, we have denoted ∇_t as the unscaled score $\partial \log p_y(y_t|f_t;\lambda)/\partial f_t$ and we let $\sup_f |\nabla_t| \in \mathbb{M}_{\Theta,\Theta}(\boldsymbol{n}, \bar{n}_\nabla)$ where $\bar{n}_\nabla$ denotes the moments of $\sup_f |\nabla_t|$.

To establish consistency, we use the following two assumptions.

Assumption 4.4. *$\exists\ \Theta^* \subseteq \mathbb{R}^4$ and $n_f > 0$ such that, for every $\hat{f}_1 \in \mathcal{F}$,*

(i) $\|s(\hat{f}_1, y_t; \cdot)\|_{n_f}^{\Theta^*} < \infty$;

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

Assumption 4.5. *$n_\ell = \min\{n_{\log \bar{g}'}, n_{\bar{p}}\} \geq 1$ and $\bar{n}_\nabla > 0$.*

Assumption 4.4 ensures the convergence of the sequence $\{\hat{f}_t\}$ to an SE limit with n_f moments on the parameter space Θ^*. As mentioned before, these conditions are similar to those imposed by Straumann and Mikosch (2006) for consistency of the QMLE of non-linear GARCH models and for example Meitz and Saikkonen (2011), who use a somewhat more restrictive analogue of (i) for consistency of the QMLE of non-linear AR-GARCH models. Assumption 4.5 ensures one bounded moment for the log-likelihood function and a uniform logarithmic moment for its derivative with respect to f. Both assumptions are stated in terms of the core structure of the score-driven model: the density of the innovations $\bar{p}$, the link function $\log \bar{g}'$, the unscaled score ∇_t, and the scaled score s_t. The number of bounded moments of $\bar{p}$, $\log \bar{g}'$, ∇_t and s_t can be easily determined as we have set out in Technical Appendix G. We illustrate the verification of these assumptions using our main example.

Main example (continued). *From the derivations around equation (3.6), we have learned that the conditions of Assumption 4.4 can be easily satisfied for an appropriate compact*

if a function $h \in \mathbb{C}^{(p,q)}$, then all cross-derivatives $h^{(i,j)}$ with $0 \leq i \leq p$, $0 \leq j \leq q$ and $i + j \leq \max\{p, q\}$ exist and are continuous *parameter space Θ^*. Namely, some compact set $\Theta^* \subseteq \{\boldsymbol{\theta} \in \mathbb{R}^4 : \beta \geq \alpha \geq 0\,,\ \omega > 0\,,\ \lambda > 0\,,\ \beta + \lambda\alpha < 1\}$ meets the requirements. For Assumption 4.5, we notice that $\bar{g}'(f_t, y_t) = f_t^{-1/2}$, and hence $n_{\log \bar{g}'} \to \infty$ given that $n_f > 0$ and $f_t \geq \omega > 0$ under the parameter constraint $\beta \geq \alpha \geq 0$ and the initialization $\hat{f}_1 \geq \omega > 0$. Using the expression*

$$\bar{p}_t = \log \frac{\Gamma(\frac{\lambda+1}{2})}{\Gamma(\frac{\lambda}{2})\sqrt{\lambda\pi}} - \tfrac{1}{2}(\lambda+1)\log\left(1 + \frac{y_t^2}{\lambda\, f_t}\right),$$

it follows immediately that $n_{\bar{p}}$ can be set arbitrarily large as long as $n_y > 0$. The condition $n_\ell \geq 1$ in Assumption 4.5 thus only requires the existence of some arbitrarily small moment $n_y > 0$ of the data y_t. Finally, since the unscaled score is given by

$$\nabla(f_t, y_t; \lambda) = \frac{(1+\lambda^{-1})y_t^2}{2f_t^2(1 + y_t^2/(\lambda f_t))} - \frac{1}{2f_t},$$

it is uniformly bounded in both $f_t \geq \omega$ and $y_t \in \mathbb{R}$, and hence, $\bar{n}_\nabla > 0$ is trivially satisfied.

Theorem 4.6 establishes the strong consistency of the MLE $\hat{\boldsymbol{\theta}}_T$. The limit log-likelihood $\ell_\infty(\cdot)$ that occurs in this theorem is defined as $\ell_\infty(\boldsymbol{\theta}) = \mathbb{E}\tilde{\ell}_t(\boldsymbol{\theta})\ \forall\ \boldsymbol{\theta} \in \Theta$, with $\tilde{\ell}_t$ denoting the contribution of the t-th observation to the likelihood function ℓ_T.

Theorem 4.6. (Consistency under possible model misspecification) *Let $\{y_t\}_{t\in\mathbb{Z}}$ be an SE sequence. Furthermore, let $\mathbb{E}|y_t|^{n_y} < \infty$ for some $n_y > 0$ for which also Assumptions 4.1, 4.2, 4.4, and 4.5 hold. Finally, let $\boldsymbol{\theta}_0 \in \Theta$ be the unique maximizer of the limit log-likelihood $\ell_\infty(\cdot)$ on the parameter space $\Theta \subseteq \Theta^*$ with Θ^* as introduced in Assumption 4.4. Then the MLE satisfies $\hat{\boldsymbol{\theta}}_T \stackrel{a.s.}{\to} \boldsymbol{\theta}_0$ as $T \to \infty$ for any filter initialization $\hat{f}_1 \in \mathcal{F}$.*

We emphasize that the proofs and results of Theorem 4.6 establish global rather than local consistency. In particular, the assumptions ensure the appropriate limiting behavior of the average log-likelihood over the entire parameter space Θ, rather than in a (possibly arbitrarily small) parameter space around the true parameter value only. This stands in sharp contrast with most of the existing literature on score models, which only delivers local asymptotic results in a neighborhood of $\boldsymbol{\theta}_0$.

Theorem 4.6 also differs from results in the existing score literature in that it establishes the strong consistency of the MLE in a possibly misspecified model setting. In particular, consistency of the MLE is obtained with respect to a pseudo-true parameter $\boldsymbol{\theta}_0 \in \Theta$ that is *assumed* to be the *unique* maximizer of the limit log-likelihood $\ell_\infty(\boldsymbol{\theta})$. This pseudo-true parameter minimizes the Kullback-Leibler divergence between the probability measure of $\{y_t\}_{t\in\mathbb{Z}}$ and the measure implied by the model. Despite the misspecification of the model, conducting inference on a pseudo-true parameter is interesting in itself. In particular, inference on pseudo-true parameters allows us to ask question about the best approximation to the data generating process (DGP). The value of this type of inference is well established in the work of Halbert White since White (1980) which focuses on the interpretation of linear misspecified approximations to nonlinear DGPs; see also White (1982), Byron and Bera (1983), Gourieroux et al. (1984), the textbook White (1994) for an extensive and detailed analysis of econometric inference under misspecification, and Gouriéroux et al. (2019) for a recent addition to this literature. Note also that this literature differs from the local-robustness literature, or the QMLE literature, which deals with small forms of model misspecification which still allows us to conduct inference on true parameters; see e.g. Newey and Steigerwald (1997) for an early discussion of the limitations of QMLE or Buja et al. (2019) for a more recent example of such efforts which links to the recent statistical and machine learning literature.

In case of misspecification, it is generally difficult to ensure the uniqueness of $\boldsymbol{\theta}_0$, so this assumption might fail. See for example Chapter 4 of Pötscher and Prucha (1997) for a discussion of this point. Luckily, uniqueness is not crucial for consistency, raising issues only if one wishes to conduct inference using standard asymptotic normality results. In case the limit criterion is maximized by a set of points Θ_0, then set consistency can be ensured without any additional assumptions. We state this result in the corollary below.

Corollary 4.7. (Set consistency under possible model misspecification) *Let $\{y_t\}_{t\in\mathbb{Z}}$ be an SE sequence. Furthermore, let $\mathbb{E}|y_t|^{n_y} < \infty$ for some $n_y \geq 0$ for which also Assumptions 4.1, 4.2, 4.4, and 4.5 hold. Finally, let Θ_0 be the set of maximizers of the limit log-likelihood $\ell_\infty(\cdot)$ on the parameter space $\Theta \subseteq \Theta^*$ with Θ^* as introduced in Assumption 4.4. Then the MLE $\hat{\boldsymbol{\theta}}_T$ satisfies $\inf_{\boldsymbol{\theta}_0\in\Theta_0} |\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0| \xrightarrow{a.s.} 0$ as $T \to \infty$ for any filter initialization $\hat{f}_1 \in \mathcal{F}$.*

This corollary ensures set consistency of the estimator towards Θ_0, and hence it ensures again that we minimize the KL divergence with respect to the true data generating process, in the limit, as T diverges to infinity. This result follows from Lemma 4.2 in Pötscher and Prucha (1997), which requires the uniform convergence of the criterion to a limit criterion with so-called "regular level sets". Luckily, the regularity of the level sets, see Pötscher and Prucha (1997, Definition 4.1), follows trivially from the compactness of Θ and the continuity of the limit criterion. All these conditions hold under the maintained assumptions, as shown in the proof of Theorem 4.6.

The results in Theorem 4.6 and Corollary 4.7 naturally require regularity conditions on the observed data $\{y_t\}_{t=1}^T \subset \{y_t\}_{t\in\mathbb{Z}}$ that is generated by an unknown data generating process. Such conditions in this general setting can only be imposed by means of direct assumption. However, under an axiom of correct specification, we can restrict the parameter space in such a way that we can *show* that the desired assumptions hold. More specifically, we can show that y_t is stationary and has n_y moments, and $\boldsymbol{\theta}_0$ is the unique maximizer of the limit log-likelihood function. In this case, the properties of the observed data $\{y_t\}_{t=1}^T$ no longer need to be *assumed*. Instead, they can be *derived* from the properties of the score-driven model under appropriate restrictions on the parameter space. By establishing 'global identification' we ensure that the limit likelihood has a unique maximum over the entire parameter space rather than only in a small neighborhood of the true parameter. The latter is typically used in most of the existing literature and achieved by studying the local properties of the information matrix at the true parameter.

To formulate our global identification result, we introduce a slightly more precise notation concerning the domains and images of the key mappings defining the score-driven model. Define the set $\mathcal{Y}_g \subseteq \mathbb{R}$ as the image of $\mathcal{F}_g$ and $\mathcal{U}$ under g, i.e., $\mathcal{Y}_g := \{g(f,u),\ (f,u) \in \mathcal{F}_g \times \mathcal{U}\}$, where $\mathcal{F}_g$ denotes the domain (for f_t) of g. Let $\mathcal{U}$ denote the common support of $p_u(\,\cdot\,;\lambda)\ \forall\ \lambda \in \Lambda$, and let $\mathcal{F}_s$ and $\mathcal{Y}_s$ denote subsets of $\mathbb{R}$ over which the map s is defined. Furthermore, statements for almost every (f.a.e.) element in a set hold with respect to Lebesgue measure. Finally, we let $g \in \mathbb{M}_{\Theta,\Theta}(\boldsymbol{n}, n_g)$ with $\boldsymbol{n} = (n_{f^u}, n_u)$, so that n_g denotes the number of bounded moments of $g(f_t, u_t)$ when u_t has n_u moments and f_t has n_{f^u} bounded moments. In practice, the resulting n_g bounded moments can be derived from the moment preservation properties laid out in the Technical Appendix.

The following two assumptions allow us to derive the appropriate properties for $\{y_t\}_{t\in\mathbb{Z}}$ and to ensure global identification of the true parameter.

Assumption 4.8. *$\exists\ \Theta_* \subseteq \mathbb{R}^4$ and $n_u \geq 0$ such that for $\mathcal{U}$, $\mathcal{Y}_g$, $\mathcal{F}_g$ and let Λ_* denote the orthogonal projection of a set $\Theta_* \subseteq \mathbb{R}^4$ onto the subspace $\mathbb{R}$ holding the static parameter λ.*

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

(ii) $\sup_{\lambda\in\Lambda_*} \mathbb{E}|u_t|^{n_u} < \infty$ and $n_g \geq n_y > 0$;

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

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

Note there is a difference between Θ^* from Assumption 4.4, and Θ_* in Assumption 4.8. The former restricts the statistical model's parameter space to establish invertibility and moments, while the latter restricts the DGP's parameter space to establish stationarity, ergodicity and moments. Conditions *(i)* and *(iii)* of Assumption 4.8 ensure that on the parameter space Θ_* the innovations u_t have non-degenerate support and $g(f,\cdot)$ is continuously differentiable and invertible with continuously differentiable derivative. Hence the conditional distribution p_y of y_t given f_t is non-degenerate and uniquely defined by the distribution of u_t. Bounded moments for y_t up to order n_y follow from moments of u_t and f_t^u via condition *(ii)*; see the main example below for an illustration of how to operate this condition. Finally, condition *(iv)* states that the static model defined by the observation equation $y_t = g(f, u_t)$ and the density $p_u(\cdot\,; \lambda)$ is identified. It requires the conditional density of y_t given $f_t = f$ to be unique for every pair (f, λ). This requirement is very intuitive: one would not extend a static model to a dynamic one if the former is not already identified.

Main example (continued). *For the Student's t scale model, the domain of u_t is always $\mathbb{R}$, which satisfies part (i) of Assumption 4.8. Parts (iii) and (iv) follow directly from the specification of the model $g(f, u) = f^{1/2}u$ and the Student's t density. Finally, as $g(f, u) = f^{1/2}u$, we can use a standard Hölder inequality to obtain $n_g = 2n_{f^u} \cdot n_u/(n_u + 2n_{f^u})$, such that part (ii) is satisfied for $n_{f^u} > 0$, $0 < n_u < \inf_{\Lambda_*} \lambda$. Note that n_{f^u} follows from Proposition 3.1, part (iii), and can be set arbitrarily high for $\boldsymbol{\theta} \in \Theta^*$, as will be explained in the discussion after Assumption 4.9.*

Assumption 4.9. *$\exists\ \Theta_* \subseteq \mathbb{R}^4$ and $n_{f^u} > 0$, such that for every $\boldsymbol{\theta} \in \Theta_*$ and every $\hat{f}_1^u \in \mathcal{F}_s$*

(i) $\|s_u(\hat{f}_1^u, u_1; \lambda)\|_{n_{f^u}} < \infty$;

(ii) $\mathbb{E}\rho_t^{n_{f^u}}(\boldsymbol{\theta}) < 1$;

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

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

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

Conditions *(i)* and *(ii)* in Assumption 4.9 ensure that on the parameter space Θ_* the true sequence $\{f_t(\boldsymbol{\theta}_0)\}$ is SE and has n_{f^u} moments by the application of Proposition 3.1. Together with condition *(iii)* in Assumption 4.8 we then obtain that the data $\{y_t\}_{t\in\mathbb{Z}}$ itself is SE and has n_y moments. The inequality stated in (4.1) in Assumption 4.9 and the assumption that $\alpha \neq 0$ together ensure that the data $\{y_t\}$ entering the update equation (1.1) render the filtered sequence $\{f_t\}$ stochastic and non-degenerate. Also, together with Assumption 4.8 part *(i)*, which ensures there is enough variation in the observed data, these two assumptions imply identification of the time-varying parameter $f_t(\boldsymbol{\theta})$. In other papers on non-linear observation-driven models, such as Straumann and Mikosch (2006) and Meitz and Saikkonen (2011), the identification restriction is formulated in a more general way and in each of their examples, they determine model-specific identification restrictions. These restrictions often include that the innovation distribution is not concentrated at two points. In the current context, such a condition is not necessary, because by Assumption 4.1 *(i)* the support of the innovations is an open set.

Next we show that our leading example satisfies the conditions for our global identification result.

Main example (continued). *The score s_u is the product of f_t^u and a term that is uniformly bounded in u_t. Hence, (i) in Assumption 4.9 is satisfied for arbitrary $n_{f^u} > 0$. Furthermore, by the linearity of s_u in f_t^u, condition (ii) of Assumption 4.9 collapses to*

$$\mathbb{E}\left|\beta - \alpha + \alpha\frac{(1+\lambda^{-1})u_t^2}{1+u_t^2/\lambda}\right|^{n_{f^u}} < 1.$$

In particular, for $n_{f^u} = 1$, we obtain the requirement $|\beta| < 1$, which together with the parameter restrictions to ensure positivity of f_t result in $1 > \beta \geq \alpha > 0$. Notice that we also require $\alpha \neq 0$ now. Larger regions can be obtained for smaller values of n_{f^u}. Notice that n_{f^u} can be set arbitrarily high for $\boldsymbol{\theta} \in \Theta^*$, so when $\beta + \lambda\alpha < 1$. In other words, Assumption 4.8 and 4.9 impose no further restrictions on the parameter space, apart from the condition that $\alpha \neq 0$, so for Θ_* we can simply take a compact subset $\Theta_* \subseteq \Theta^* \setminus \{\boldsymbol{\theta} \in \mathbb{R}^4 : \alpha = 0\}$.

Theorem 4.10 (Global Identification for correctly specified models)**.** *Let Assumptions 4.1, 4.2, 4.4, 4.5, 4.8, and 4.9 hold and let the observed data be a subset of the realized path of a stochastic process $\{y_t\}_{t\in\mathbb{Z}}$ generated by a score-driven model under $\boldsymbol{\theta}_0 \in \Theta$. Then $Q_\infty(\boldsymbol{\theta}_0) \equiv \mathbb{E}_{\boldsymbol{\theta}_0}\ell_t(\boldsymbol{\theta}_0) > \mathbb{E}_{\boldsymbol{\theta}_0}\ell_t(\boldsymbol{\theta}) \equiv Q_\infty(\boldsymbol{\theta})\ \forall\ \boldsymbol{\theta} \in \Theta : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$.*

The axiom of correct specification thus leads to the global identification result in Theorem 4.10. We can use this to establish consistency of the MLE to the true (rather than pseudo-true) parameter value if the model is correctly specified. This is summarized in the following corollary.

Corollary 4.11. (Consistency for correctly specified models) *Let Assumptions 4.1, 4.2, 4.4, 4.5, 4.8, and 4.9 hold and $\{y_t\}_{t\in\mathbb{Z}} = \{y_t(\boldsymbol{\theta}_0)\}_{t\in\mathbb{Z}}$ with $\boldsymbol{\theta}_0 \in \Theta$, where $\Theta \subseteq \Theta^* \cap \Theta_*$ with Θ^* and Θ_* defined in Assumptions 4.4, 4.8 and 4.9. Then the MLE $\hat{\boldsymbol{\theta}}_T$ satisfies $\hat{\boldsymbol{\theta}}_T \stackrel{a.s.}{\to} \boldsymbol{\theta}_0$ as $T \to \infty$ for every $\hat{f}_1 \in \mathcal{F}$.*

The consistency region $\Theta^* \cap \Theta_*$ under correct specification is a subset of the consistency region Θ^* for the misspecified setting. In Theorem 4.6, we namely assume that the data is SE and that the true parameter is identified, while in Corollary 4.11 we do not make these assumptions directly. On the parameter space Θ^*, the filtered sequence $\{\hat{f}_t\}$ converges uniformly to an SE limit with a certain number of moments, but this is no longer enough for consistency without the direct assumption of SE data and identification of the true parameter. The parameter space also has to be (further) restricted to Θ_* to ensure that the score-driven data generating process is identified and generates SE data with the appropriate number of moments.

To establish asymptotic normality of the MLE, we impose an assumption that delivers $2+\delta$ moments for some small positive δ for the first derivative of the log-likelihood function, and 1 moment for the second derivative. We make use once again of our notation for moment preserving maps. In particular, quantities like $n_{\bar{p}}^{\lambda}$ denote the number of bounded moments of the derivative of $\bar{p}$ with respect to λ and quantities like $\bar{n}_{\bar{p}}^{\lambda f}$ denote the number of bounded moments of the supremum over f_t of the cross derivative with respect to f and λ. We also let $n_{f_{\boldsymbol{\theta}}}$ and $n_{f_{\boldsymbol{\theta\theta}}}$ be defined as in Proposition 3.5.

Assumption 4.12. *$\exists\ \Theta_*^* \subseteq \mathbb{R}^4$ such that $n^\star > 0$, $n_{\ell''} \geq 1$ and $n_{\ell'} > 2$, with*

$$\begin{aligned} n^\star \;&=\; \min\big\{\bar{n}_\nabla \;,\; \bar{n}_\nabla^{f} \;,\; n_\nabla^{\lambda} \;,\; \bar{n}_\nabla^{ff} \;,\; \bar{n}_\nabla^{\lambda f} \;,\; \bar{n}_{\bar{p}}^{\lambda f} \;,\; \bar{n}_{\bar{p}}^{\lambda\lambda f} \;,\; n_s \;,\; n_s^{\lambda} \;, \\ &\qquad\quad \bar{n}_s^{f} \;,\; \bar{n}_s^{\lambda f} \;,\; \bar{n}_s^{ff} \;,\; n_s^{\lambda\lambda} \;,\; \bar{n}_s^{\lambda\lambda f} \;,\; \bar{n}_s^{\lambda ff} \;,\; \bar{n}_s^{fff}\big\} \;, \end{aligned} \tag{4.2}$$

$$n_{\ell'} \;=\; \min\left\{ n_{\bar{p}}^{\lambda} \;,\; \frac{n_\nabla n_{f_{\boldsymbol{\theta}}}}{n_\nabla + n_{f_{\boldsymbol{\theta}}}} \right\}, \tag{4.3}$$

$$n_{\ell''} \;=\; \min\left\{ n_{\bar{p}}^{\lambda\lambda} \;,\; \frac{n_\nabla n_{f_{\boldsymbol{\theta\theta}}}}{n_\nabla + n_{f_{\boldsymbol{\theta\theta}}}} \;,\; \frac{n_\nabla^{\lambda} n_{f_{\boldsymbol{\theta}}}}{n_\nabla^{\lambda} + n_{f_{\boldsymbol{\theta}}}} \;,\; \frac{n_\nabla^{f} n_{f_{\boldsymbol{\theta}}}}{2n_\nabla^{f} + n_{f_{\boldsymbol{\theta}}}} \right\}. \tag{4.4}$$

We introduce the new set Θ_*^*, because the parameter restrictions imposed by the parameter space Θ^* are not always strong enough to ensure the existence of all the moments in Proposition 3.3. So for asymptotic normality, we need to further restrict the parameter space because we need these additional moment conditions to hold.

Similar to the moment conditions in Proposition 3.3, the moment conditions in Assumption 4.12 relate directly to low-level (primitive) elements of the model. The expressions in (4.2), (4.3) and (4.4) follow directly from the formulas for the derivatives of the log-likelihood with respect to $\boldsymbol{\theta}$. Having $n_{\ell'} > 2$ facilitates the application of a central limit theorem to the score. Similarly, $n_{\ell''} \geq 1$ allows us to use a uniform law of large numbers for the Hessian. Finally, the condition $n^\star > 0$ is designed to ensure that the moment conditions of Proposition 3.5 are satisfied and the e.a.s. convergence of the filter $\hat{f}_t$ to its stationary limit is appropriately reflected in the convergence of both the score and the Hessian.

In any case, if one favors simplicity at the cost of some generality, then the expressions for $n_{\ell'}$ and $n_{\ell''}$ can be easily simplified to a single moment condition as stated in the following remark.

Remark 4.13. Let n denote the lowest of the primitive derivative moment numbers $n_{\bar{p}}^{\lambda}$, n_{∇}, etc. Then $n > 4$ implies $n_{\ell'} > 2$ and $n_{\ell''} \geq 1$.

It can be easier, however, to check the moment conditions formulated in Assumption 4.12 directly rather than the simplified conditions in Remark 4.13. We can illustrate this point using our main example.

Main example (continued). *For the Student's t scale model, a number of derivative functions need to be computed. These can be found in the Technical Appendix. Many of these are uniformly bounded functions. In particular, we have* $\bar{n}_{\nabla}, \bar{n}_{\nabla}^{f}, \bar{n}_{\nabla}^{\lambda}, \bar{n}_{\nabla}^{\lambda f}, \bar{n}_{\nabla}^{ff}, \bar{n}_{\bar{p}}^{\lambda f}, n_{\bar{p}}^{\lambda\lambda}, \bar{n}_{\bar{p}}^{\lambda\lambda f} \to \infty$*. Also recall that we argued that all moments necessary for Proposition 3.5 exist. Furthermore,* $n_{\bar{p}}^{\lambda} \leq n_y/\delta$ *for some (small)* $\delta > 0$*. Therefore, if some finite moment of* y_t *exists, we can set* $n_{\bar{p}}^{\lambda}$ *arbitrarily large. As a result,* $n^{\star} > 0$, $n_{\ell'} \leq \min\{n_y/\delta', n_{f_{\boldsymbol{\theta}}}\}$ *for arbitrary* $\delta' > 0$*, and* $n_{\ell''} \leq \min\{n_{f_{\boldsymbol{\theta}\boldsymbol{\theta}}}, n_{f_{\boldsymbol{\theta}}}, \frac{1}{2}n_{f_{\boldsymbol{\theta}}}\}$*. We have derived earlier that* $n_{f_{\boldsymbol{\theta}\boldsymbol{\theta}}} < n_{f_{\boldsymbol{\theta}}}/2$*, such that* $n_{\ell'} > 2$ *and* $n_{\ell''} \geq 1$ *imply that we need* $n_{f_{\boldsymbol{\theta}}} > 2$*. If the contraction condition is met over the entire parameter space, then as also shown earlier we can set* $n_{f_{\boldsymbol{\theta}}}$ *arbitrarily high and thus satisfy Assumption 4.12. This condition is met on the parameter space* Θ^**, so we need no additional restriction and we can set* $\Theta^*_* = \Theta^*$*.*

In well-specified models, the asymptotic normality of the MLE is obtained by applying a central limit theorem (CLT) for SE martingale difference sequences to the ML score, that is the derivative of the log-likelihood function $\ell_T(\boldsymbol{\theta}, \hat{f}_1)$ with respect to $\boldsymbol{\theta}$ and evaluated at the MLE. As noted in White (1994), in the presence of dynamic misspecification, the ML score generally fails to be a martingale difference sequence even at the pseudo-true parameter. As a result, stricter conditions are required to obtain a central limit theorem that allows for some temporal dependence in the ML score.

Below we use the property of near epoch dependence (NED) to obtain a CLT for the ML score. In particular, we use the uniform filter contraction in Assumption 4.4 to ensure that the filter is NED whenever the data is NED. Furthermore, in Assumption 4.14 below, we impose sufficient conditions for the ML score to be Lipschitz continuous on the data as well as on the filter and its derivative. This assumption is designed to guarantee that the ML score inherits the NED property from the data and the filter. The conditions of Assumption 4.14 can be weakened in many ways; see, for example, Davidson (1994) and Pötscher and Prucha (1997) for a discussion of alternative conditions. Here the Lipschitz continuity condition allows us to keep the asymptotic normality results clear and simple.

Assumption 4.14. $\partial\bar{p}_t/\partial f$ *and* $\partial \log \bar{g}_t'/\partial f$ *are uniformly bounded random variables and* $\partial\bar{p}_t/\partial f$*,* $\partial \log \bar{g}_t'/\partial f$ *and* $\partial\bar{p}_t/\partial\lambda$ *are a.s. Lipschitz continuous in* (y_t, f_t)*.*

Main example (continued). *Using the Student's t scale model, we have already seen that $f_t \geq \omega > 0$ for all t. The relevant derivative of $\bar{p}_t$ equals f_t^{-1} times a uniformly bounded function of y_t^2/f_t, which obviously results in a uniformly bounded function. Also $\partial \log \bar{g}_t'/\partial f = 0.5 f_t^{-1}$ is trivially uniformly bounded. Furthermore, $\partial \bar{p}_t/\partial f$, $\partial \log \bar{g}_t'/\partial f$ and $\partial \bar{p}_t/\partial \lambda$ are Lipschitz continuous in (y_t, f_t), because their first derivatives with respect to y_t and f_t are bounded. Hence Assumption 4.14 holds for the leading example.*

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

Theorem 4.15. (Asymptotic normality under possible model misspecification) *Let $\{y_t\}_{t\in\mathbb{Z}}$ be SE and NED of size -1 on a strongly mixing process of size $-\delta/(1-\delta)$ for some $\delta > 2$. Furthermore, let $\mathbb{E}|y_t|^{n_y} < \infty$ for some $n_y \geq 0$ for which also Assumptions 4.1, 4.2, 4.4, 4.5, 4.12 and 4.14 are satisfied. Finally, let $\boldsymbol{\theta}_0 \in \mathrm{int}(\Theta)$ be the unique maximizer of $\ell_\infty(\boldsymbol{\theta})$ on Θ, where $\Theta \subseteq \Theta^* \cap \Theta_*^*$ with Θ^* and Θ_*^* as defined in Assumptions 4.4 and 4.12 and let $\mathbb{E}\tilde{\ell}_t''(\boldsymbol{\theta}_0)$ be non-singular. Then, for every $\hat{f}_1 \in \mathcal{F}$, the MLE $\hat{\boldsymbol{\theta}}_T(\hat{f}_1)$ satisfies*

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

where $\mathcal{I} := -\mathbb{E}\tilde{\ell}_t''(\boldsymbol{\theta}_0)$ is the Fisher information matrix, $\tilde{\ell}_t(\boldsymbol{\theta}_0)$ denotes the log-likelihood contribution of the t-th observation evaluated at $\boldsymbol{\theta}_0$, and

$$\mathcal{J}(\boldsymbol{\theta}_0) := \lim_{T\to\infty} T^{-1}\mathbb{E}\Big(\sum_{t=1}^{T} \tilde{\ell}_t'(\boldsymbol{\theta}_0)\Big)\Big(\sum_{t=1}^{T} \tilde{\ell}_t'(\boldsymbol{\theta}_0)^\top\Big).$$

When the model is correctly specified, the ML score can be shown to be a martingale difference sequence at the true parameter value. Hence we no longer need the assumption that the data is NED. Also, we can drop Assumption 4.14, which was used to ensure that the ML score was NED. In general we are presented with a trade-off between the assumption of correct specification combined with weaker additional assumptions, versus the stricter NED conditions without the assumption of correct specification. Apart from this trade-off, the proof of asymptotic normality is the same in both cases. The following theorem states the asymptotic normality result for the MLE in the context of a correctly specified model.

Theorem 4.16. (Asymptotic normality under correct specification) *Let Assumptions 4.1, 4.2, 4.4, 4.5, 4.8, 4.9, and 4.12 hold and assume $\{y_t\}_{t\in\mathbb{Z}}$ is a random sequence generated by a score-driven model under some $\boldsymbol{\theta}_0 \in \mathrm{int}(\Theta)$ where $\Theta \subseteq \Theta^* \cap \Theta_* \cap \Theta_*^*$ with Θ^*, Θ_* and Θ_*^* defined in Assumptions 4.4, 4.8, 4.9, and 4.12 and let $\mathbb{E}\tilde{\ell}_t''(\boldsymbol{\theta}_0)$ be regular in the sense of Rothenberg (1971). Then, for every $\hat{f}_1 \in \mathcal{F}$, the MLE $\hat{\boldsymbol{\theta}}_T(\hat{f}_1)$ satisfies*

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

where $\mathcal{I}$ is the Fisher information matrix as defined in Theorem 4.15.

Theorem 4.16 does not have a separate n_y-moment condition like Theorems 4.6 and 4.15. This stems from the fact that under correct specification the moment conditions for y_t are implied by the moment conditions on the data generating process, such as the moment conditions on u_t and $g(f_t, y_t)$ in Assumptions 4.8 and 4.9.

Main example (continued). *To verify the conditions of Theorem 4.16 for the main example, we have already shown that Assumption 4.12 requires $n_f > 2$ and that an arbitrarily small moment $n_y > 0$ of y_t exists. Using the derivations below Proposition 3.3, we showed that the condition $n_f > 2$ is met if the contraction condition (3.6) is satisfied. Furthermore, we already showed in the exposition after Assumption 4.8 that the condition $n_y > 0$ is met if $\inf_\Lambda \lambda = \underline{\lambda} > 0$ such that an arbitrarily small moment exists for u_t and if $n_{f^u} > 0$, which holds under Assumption 4.9.*

5 Empirical Illustration

The theorems and corollaries derived in the previous section establish the existence, strong consistency, global identification, and asymptotic normality of the MLE for a general class of score-driven models under correct and incorrect model specification. In this section, we make use of a practical example to provide some further intuition for the main assumptions and results next to the leading example dealt with throughout the main text.

The Student's t Location Model

Consider the score-driven Student's t location model proposed by Harvey and Luati (2014). The observation equation of the model is given by

$$y_t = f_t + u_t \ , \quad f_{t+1} = \omega + \alpha\, w_t\,(y_t - f_t) + \beta f_t \ , \quad w_t = (1 + \nu^{-1} e^{-2\kappa}\,(y_t - f_t)^2)^{-1}, \quad (5.1)$$

where we use a scaling function $S(f_t; \lambda) = (1 + \nu^{-1})^{-1} e^{2\kappa}$ proportional to the inverse conditional Fisher information. So in the notation of (1.1), we have $g(f_t, u_t) = f_t + u_t$, which is strictly increasing in u_t, and we impose that u_t has a Student's t-density p_u with degrees of freedom parameter $\nu > 0$ and scale parameter $\exp(\kappa)$. So $\lambda = (\nu, \kappa)$ is two-dimensional. As argued before, all results continue to hold for multivariate λ. Clearly, the inverse link function and its derivative with respect to y_t are given by $\bar{g}_t = g^{-1}(f_t, y_t) = y_t - f_t$ and $\bar{g}'_t = 1$.

Consistency

If the model is well specified, then we can show consistency of the MLE on a compact Θ by demonstrating that the assumptions of Corollary 4.11 hold. It is straightforward to see that assumptions 4.1 and 4.2 hold for this model. Next, we note that Assumption 4.4 holds on some $\Theta^* \subseteq \mathbb{R}^5$ and for some $n_f > 0$, which ensures the uniform invertibility of the filter. In particular, condition *(i)* holds for all compact sets Θ^* and every $n_f > 0$, because $s(\hat{f}_1, y_t; \cdot)$ is uniformly bounded in y_t for any given $\hat{f}_1$. For condition *(ii)*, it can be shown that for any λ and y_t, the derivative

$$\frac{\partial s(f^*, y_t; \lambda)}{\partial f} = \frac{\nu^{-1} e^{-2\kappa}(y_t - f^*)^2 - 1}{(1 + \nu^{-1} e^{-2\kappa}(y_t - f^*)^2)^2},$$

is bounded between -1 and $1/8$ (these values are attained at $y_t - f^* = 0$ and $y_t - f^* = \pm \exp(\kappa)\sqrt{3\nu}$ respectively). It follows that

$$\sup_{(f^*, y, \boldsymbol{\theta}) \in \mathcal{F} \times \mathcal{Y} \times \Theta^*} \left| \beta + \alpha \frac{\partial s(f^*, y; \lambda)}{\partial f} \right| \le \sup_{\boldsymbol{\theta} \in \Theta^*} \max \left\{ |\beta - \alpha|, \left| \beta + \frac{1}{8}\alpha \right| \right\}.$$

Therefore, condition *(ii)* of Assumption 4.4 holds for any compact set $\Theta^* \subseteq \mathbb{R}^5$ for which $\max\{|\beta - \alpha|, |\beta + \frac{1}{8}\alpha|\} < 1$ and $\nu > 0$ for every $\boldsymbol{\theta} \in \Theta^*$. The region of the (α, β)-plane where the filter is invertible is represented by the hatched area in Figure 1. So, Assumption 4.4 holds for any compact Θ^* with all pairs (α, β) in the interior of this area (and $\nu > 0$), and the filter is invertible uniformly over Θ^*.

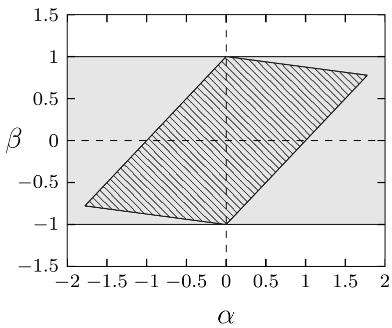

Figure 1: The stationarity region (the gray shaded area) and the invertibility region (the hatched area) of α and β for the score-driven Student's t location model. These areas contain the pairs (α, β) for which $|\beta| < 1$ and $\max\{|\beta - \alpha|, |\beta + \frac{1}{8}\alpha|\} < 1$, respectively.

For Assumption 4.5, recall that $\bar{g}'_t = 1$, so we only need to show that $n_{\bar{p}} \ge 1$ and $\bar{n}_\nabla > 0$. For the former, consider the expression

$$\bar{p}_t = \log \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\nu\pi}e^\kappa} - \tfrac{1}{2}(\nu + 1) \log \left(1 + \nu^{-1} e^{-2\kappa}(y_t - f_t)^2\right).$$

It follows directly that $n_{\bar{p}}$ can be set arbitrarily high if $\mathbb{E}|y_t - f_t|^n < \infty$ for some $n > 0$. By the C_r-inequality in Loève (1977, p.157), we have that $\mathbb{E}|y_t - f_t|^n \leq c\mathbb{E}|y_t|^n + c\mathbb{E}|f_t|^n$ for some $c > 0$, so $n_{\bar{p}} > 1$ only requires $n_f > 0$ and $n_y > 0$, where the former condition holds by Assumption 4.4 and the latter is either assumed directly (in Theorem 4.6) or indirectly (in Corollary 4.11). Also, for $\bar{n}_\nabla > 0$, we can use that the unscaled score is given by

$$\nabla(f_t, y_t; \nu) = (1 + \nu^{-1}) \frac{e^{-2\kappa}(y_t - f_t)}{1 + \nu^{-1}e^{-2\kappa}(y_t - f_t)^2},$$

which is uniformly bounded in $(f_t, y_t) \in \mathbb{R}^2$. This implies that $\bar{n}_\nabla > 0$ is trivially satisfied.

If the model is *well specified*, the MLE is consistent for the true parameter $\boldsymbol{\theta}_0$ as long as it generates appropriately behaved data as a DGP; see Technical Appendix E for relevant details on the verification of the DGP assumptions 4.8 and 4.9. In contrast, if the model is *misspecified*, then the MLE is set-consistent with respect to the set of pseudo-true parameters, as long as observed data is stationary, by Corollary 4.7. If there exists a unique pseudo-true parameter $\boldsymbol{\theta}_0$, then the MLE is consistent for $\boldsymbol{\theta}_0$ by Theorem 4.6.

Asymptotic normality

Asymptotic normality can be obtained by verifying Assumption 4.12 (or the simpler condition in Remark 4.13) as well as Assumption 4.14. Assumption 4.12 requires the existence of certain moments. All quantities under consideration are uniformly bounded in both $f_t \in \mathbb{R}$ and $y_t \in \mathbb{R}$ as long as $\nu > \underline{\nu} > 0$, except for $\partial\bar{p}_t/\partial\nu$. This derivative namely consists of uniformly bounded terms and the term $\log(1 + \nu^{-1}e^{-2\kappa}(y_t - f_t)^2)$, which is not uniformly bounded. However, we know that this term has bounded moments of any order as long as $n_y > 0$ and $n_f > 0$, which holds by previous assumptions. Hence, Assumption 4.12 holds under the current parameter restrictions because $n^\star$, $n_{\ell''}$ and $n_{\ell'}$ can all be set arbitrarily high. Therefore, we can choose $\Theta^*_* = \Theta^*$, because the moment conditions in Assumption 4.12 does not impose any additional restrictions on the parameter space. Finally, Assumption 4.14 holds, because it can be seen straightforwardly that $\partial\bar{p}_t/\partial f$ is uniformly bounded, $\partial\bar{p}_t/\partial f$ and $\partial\bar{p}_t/\partial\lambda$ are a.s. Lipschitz continuous in (y_t, f_t) and $\partial \log \bar{g}'_t/\partial f = 0$.

Under *correct specification*, and the maintained assumptions, we have by Theorem 4.16 that the MLE $\hat{\boldsymbol{\theta}}_T$ is consistent for the true parameter $\boldsymbol{\theta}_0$ and asymptotically normal with an asymptotic variance given by the inverse information matrix, in case $\mathbb{E}\tilde{\ell}''_t(\boldsymbol{\theta}_0)$ is regular in the sense of Rothenberg (1971). If the model is misspecified, then by Theorem 4.15 the MLE is consistent for a pseudo-true parameter $\boldsymbol{\theta}_0$ and asymptotically normal with an asymptotic

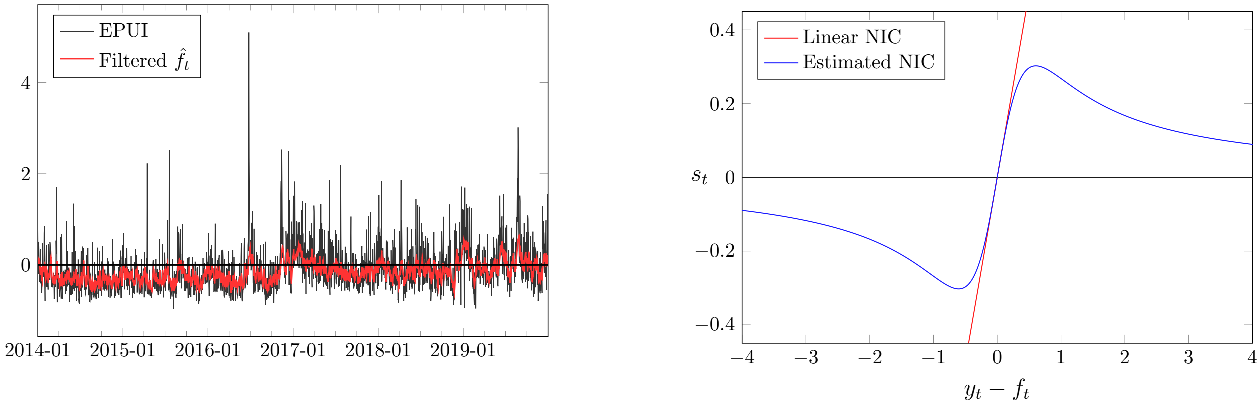

Figure 2: Normalized daily EPUI data from 2014 to 2019 (left panel) and News Impact Curve (NIC) of a linear and estimated filter (right panel). The left panel also holds the filtered $\hat{f}_t$ using the estimated Student's t score-driven filter.

variance given by a sandwich formula, whenever the assumptions of Theorem 4.6 hold, the unique maximizer $\boldsymbol{\theta}_0$ of $\ell_\infty(\boldsymbol{\theta})$ on Θ lies in the interior of Θ, and $\mathbb{E}\tilde{\ell}_t''(\boldsymbol{\theta}_0)$ is non-singular.

Application to EPUI data

To demonstrate how the model above could be used in practice, we apply it to the daily Economic Policy Uncertainty Index (EPUI) of the United States[2]. The EPUI has been shown to successfully proxy changes in policy-related economic uncertainty, see Baker et al. (2015). Figure 2 plots the data. We use the $T = 2191$ daily observations from 2014 until 2019.

The figure shows that the EPUI is noisy and occasionally displays large outliers, such as the June 2016 spike, which is probably due to the Brexit referendum outcome. After spikes like this, the EPUI usually quickly returns to its mean value. Therefore, the Student's t location model could be a suitable model for this data, as it downweighs large observations in the construction of the filtered location. This is most clearly seen when looking at the news impact curves (NIC) in the right panel of Figure 2.

The maximum likelihood estimates are given in Table 1. Notice that the estimated degrees of freedom parameter $\hat{\nu}_T$ has a low value of around 4, which implies a fat-tailed innovation distribution and more importantly shows that the estimated model has a filter which is robust to large outliers. The filtered locations $\{\hat{f}_t\}_{t=1}^T$ are plotted in the left panel

[2]Baker, Scott R., Bloom, Nick and Davis, Stephen J., Economic Policy Uncertainty Index for United States [USEPUINDXD], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/USEPUINDXD. We standardize the data by subtracting the sample mean over the pre-sample period 1985–2008, and dividing by 100.

Table 1: MLE estimates for the EPUI data from Figure 2 and the score-driven Student's t location model (5.1). Standard errors in parentheses.

ω	α	β	ν	κ
Full model				
-0.010	0.387	0.934	4.019	-1.198
(0.003)	(0.043)	(0.016)	(0.336)	(0.025)
Restricted model				
-0.163	0.576		4.140	-1.126
(0.018)	(0.039)		(0.479)	(0.034)

of Figure 2. The robustness of the filter is clearly visible in this graph. Incidental spikes hardly move the filter.

The parameter estimates all lie within $\Theta^* \cap \Theta_* \cap \Theta_*^*$ from Figure 1. It follows from Corollary 4.11 and Theorem 4.16 that the MLE is consistent and asymptotically normal with covariance matrix $\mathcal{I}^{-1}$ under the assumption that the model is correctly specified. The standard errors shown in Table 1 are based on the assumption of correct specification,

As a simple illustration of hypothesis testing, we consider testing the null hypothesis that $\omega = 0$. Under the assumption of correct specification, the unconditional mean of the true time-varying parameter f_t is $\omega/(1-\beta)$. Therefore, testing $H_0 : \omega = 0$ amounts to testing whether the unconditional expectations of both the time-varying location parameter f_t and the observations y_t are equal to zero. The t-statistic of $\hat{\omega}_T$ equals -3.130 and shows that the null-hypothesis is rejected at a 1% significance level indicating that the expected value of the EPUI from 2014 up until 2019 is significantly different from its pre-sample (1985–2008) average, as we use the demeaned data in our analysis.

As a second illustration, we consider estimating a restricted version of our score-driven location model by imposing $\beta = 0$. The results are provided in the lower panel of Table 1. The restriction $\beta = 0$ causes the filtered location to be less flexible compared to the unrestricted setting. Let us assume that the model is incorrectly specified. From Corollary 4.7 we know that the MLE is set-consistent to the set of pseudo-true parameters within the compact $\Theta \subseteq \Theta^* \cap \Theta_*^*$, as long as we assume that the data comes from an SE sequence which has a small bounded moment. The set of pseudo-true parameters is the collection of parameter values that minimize the limit Kullback-Leibler divergence with respect to the true distribution of the data.

If we additionally assume that the pseudo-true parameter $\boldsymbol{\theta}_0$ is unique and lies in the interior of the parameter space Θ, that $\{y_t\}_{t\in\mathbb{Z}}$ is NED of size -1 on a strongly mixing process of size $-\delta/(1-\delta)$ for some $\delta > 2$ and that $-\mathbb{E}\tilde{\ell}_t''(\boldsymbol{\theta}_0)$ is invertible, the MLE is asymptotically normal by Theorem 4.15. The corresponding standard errors can be found in Table 1 again. By Theorem 4.15 we have that the true asymptotic covariance matrix is given by $\mathcal{I}^{-1}\mathcal{J}\mathcal{I}^{-1}$, which we estimate by its sample counterpart $\hat{\mathcal{I}}_T^{-1}\hat{\mathcal{J}}_T\hat{\mathcal{I}}_T^{-1}$.

Using the estimated parameters and their standard errors, we can now test the null-hypothesis $H_0 : \alpha = 0$. The t-statistic of $0.576/0.039 \approx 14.77$ clearly shows that the null hypothesis can be rejected at any sensible significance level. This means that there is statistical evidence that the 'best approximating model' in terms of KL divergence for these data has a time-varying conditional mean. In other words, we reject the null hypothesis that the conditional mean of the EPUI data is constant over time.

The test examples considered in this section are acknowledgedly simple, but already show that the theory developed in this paper can be used for interesting test formulations relating to the best approximating model in a KL sense, even if this model is misspecified. Further tests would include (i) tests for leverage effects and asymmetry parameters in score-driven volatility models with asymmetric Gaussian or asymmetric Student's t densities for the innovations (Lucas et al., 2014, 2017; Harvey and Sucarrat, 2014); (ii) testing for mixture parameters in score models with mixture distributions (Catania, 2020); and (iii) testing for parameter significance of explanatory variables in spatial regression models with time-varying parameters (Blasques et al., 2016; Catania and Billé, 2017).

We finally note that when misspecification affects only some elements of the model, then one could still potentially obtain consistency to a true parameter using QMLE results. Some of these results also directly apply to the setting of score-driven models, such as the GARCH and the linear location model. However, the study of such small forms of misspecification is not the focus of the current paper.

6 Conclusions

We have developed an asymptotic distribution theory for the class of score-driven time-varying parameter models. Despite a wide range of newly developed models using the score-driven approach, a theoretical basis has been missing. We have aimed in this study to make a substantial step forward. In particular, we have developed a global asymptotic theory for the maximum likelihood estimator for score-driven time series models as introduced by Creal et al. (2011, 2013) and Harvey (2013). Our theorems are global in nature and are based on primitive, low-level conditions stated in terms of functions that make up the core of the score-driven model. We also state conditions under which the score-driven model is invertible. In contrast to the existing literature on score-driven models, we do not need to rely on the empirically untenable assumption that the starting value $\hat{f}_1$ is both random and observed. For the case of correctly specified models, we have been able to establish a global identification result that holds under weak conditions. We believe that the presented results establish a proper foundation for the use of the score function in observation-driven models and for maximum likelihood estimation and hypothesis testing.

References

Andres, P. and A. C. Harvey (2012). The Dyanamic Location/Scale Model: with applications to intra-day financial data. Cambridge Working Papers in Economics 1240, Faculty of Economics, University of Cambridge.

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

Babii, A., X. Chen, and E. Ghysels (2019). Commercial and residential mortgage defaults: Spatial dependence with frailty. *Journal of econometrics 212*(1), 47–77.

Baker, S. R., N. Bloom, and S. J. Davis (2015). Measuring economic policy uncertainty. Working Paper 21633, National Bureau of Economic Research.

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

Blasques, F., P. Gorgi, S. J. Koopman, and O. Wintenberger (2018). Feasible invertibility conditions and maximum likelihood estimation for observation-driven models. *Electronic Journal of Statistics 12*(1), 1019–1052.

Blasques, F., S. J. Koopman, and A. Lucas (2014). Stationarity and ergodicity conditions for generalized autoregressive score processes. *Electronic Journal of Statistics 8*, 1088–1112.

Blasques, F., S. J. Koopman, and A. Lucas (2015). Information theoretic optimality of observation driven time series models for continuous responses. *Biometrika 102*(2), 325–343.

Blasques, F., S. J. Koopman, A. Lucas, and J. Schaumburg (2016). Spillover dynamics for systemic risk measurement using spatial financial time series models. *Journal of Econometrics 195*(2), 211–223.

Blasques, F., A. Lucas, and E. Silde (2018). A stochastic recurrence equations approach for score driven correlation models. *Econometric Reviews 37*(2), 166–181.

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.

Buja, A., L. Brown, R. Berk, E. George, E. Pitkin, M. Traskin, K. Zhang, and L. Zhao (2019, 11). Models as approximations i: Consequences illustrated with linear regression. *Statist. Sci. 34*(4), 523–544.

Byron, R. P. and A. K. Bera (1983). Least squares approximations to unknown regression functions: A comment. *International Economic Review 24*(1), 255–260.

Caivano, M. and A. C. Harvey (2014). Time-series models with an EGB2 conditional distribution. *Journal of Time Series Analysis 35*(6), 558–571.

Catania, L. (2020). Dynamic adaptive mixture models with an application to volatility and risk. *Journal of Financial Econometrics*, in press.

Catania, L. and A. G. Billé (2017). Dynamic spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Applied Econometrics 32*(6), 1178–1196.

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

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

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

Creal, D., B. Schwaab, S. J. Koopman, and A. Lucas (2014). Observation driven mixed-measurement dynamic factor models. *Review of Economics and Statistics 96*(5), 898–915.

Davidson, J. (1992). A Central Limit Theorem for Globally Nonstationary Near-Epoch Dependent Functions of Mixing Processes. *Econometric Theory 8*(03), 313–329.

Davidson, J. (1993). The Central Limit Theorem for Globally Nonstationary Near-Epoch Dependent Functions of Mixing Processes: The Asymptotically Degenerate Case. *Econometric Theory 9*(03), 402–412.

Davidson, J. (1994). *Stochastic Limit Theory.* Advanced Texts in Econometrics. Oxford University Press.

De Lira Salvatierra, I. and A. J. Patton (2015). Dynamic copula models and high frequency data. *Journal of Empirical Finance 30*, 120–135.

Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica 50*, 987–1008.

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

Engle, R. F. and J. R. Russell (1998). Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica 66*(5), 1127–1162.

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

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

Gourieroux, C., A. Monfort, and A. Trognon (1984). Pseudo maximum likelihood methods: Theory. *Econometrica 52*(3), 681–700.

Gouriéroux, C., A. Monfort, and J.-M. Zakoïan (2019). Consistent pseudo-maximum likelihood estimators and groups of transformations. *Econometrica 87*(1), 327–345.

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

Hafner, C. and A. Preminger (2009). On asymptotic theory for multivariate garch models. *Journal of Multivariate Analysis 100*(9), 2044–2054.

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

Harvey, A. C. and R. Lange (2015a). Modeling the Interactions Between Volatility and Returns. Cambridge Working Papers in Economics 1518, Faculty of Economics, University of Cambridge.

Harvey, A. C. and R. Lange (2015b). Volatility Modeling with a Generalized t-Distribution. Cambridge Working Papers in Economics 1517, Faculty of Economics, University of Cambridge.

Harvey, A. C. and A. Luati (2014). Filtering with heavy tails. *Journal of the American Statistical Association 109*(507), 1112–1122.

Harvey, A. C. and G. Sucarrat (2014). EGARCH models with fat tails, skewness and leverage. *Computational Statistics & Data Analysis 76*(C), 320–338.

Janus, P., S. J. Koopman, and A. Lucas (2014). Long memory dynamics for multivariate dependence under heavy tails. *Journal of Empirical Finance 29*, 187–206.

Koopman, S. J., R. Lit, A. Lucas, and A. Opschoor (2018). Dynamic discrete copula models for high-frequency stock price changes. *Journal of Applied Econometrics 33*(7), 966–985.

Koopman, S. J., A. Lucas, and M. Scharth (2016). Predicting time-varying parameters with parameter-driven and observation-driven models. *Review of Economics and Statistics 98*, 97–110.

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

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

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

Lucas, A., B. Schwaab, and X. Zhang (2017). Modeling financial sector joint tail risk in the euro area. *Journal of Applied Econometrics 32*(1), 171–191.

Massacci, D. (2016). Tail risk dynamics in stock returns: Links to the macroeconomy and global markets connectedness. *Management Science 63*(9), 3072–3089.

Meitz, M. and P. Saikkonen (2011). Parameter estimation in nonlinear AR–GARCH models. *Econometric Theory 27*(6), 1236–1278.

Newey, W. and D. Steigerwald (1997). Asymptotic bias for quasi-maximum-likelihood estimators in conditional heteroskedasticity models. *Econometrica 65*(3), 587–599.

Oh, D. H. and A. Patton (2018). Time-varying systemic risk: evidence from a dynamic copula model of cds spreads. *Journal of Business and Economic Statistics 36*(2), 181–195.

Opschoor, A., P. Janus, A. Lucas, and D. Van Dijk (2018). New HEAVY models for fat-tailed realized covariances and returns. *Journal of Business and Economic Statistics 36*(4), 643–657.

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

Pötscher, B. M. and I. R. Prucha (1997). *Dynamic nonlinear econometric models: Asymptotic theory.* Springer.

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

Rothenberg, T. J. (1971). Identification in parametric models. *Econometrica 39*(3), 577–91.

Ryoko, I. (2016). Asymptotic theory for beta-t-GARCH. Cambridge Working Papers in Economics 1607, Faculty of Economics, University of Cambridge.

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

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

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

White, H. (1980). Using least squares to approximate unknown regression functions. *International Economic Review 21*(1), 149–170.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica 50*(1), 1–25.

White, H. (1994). *Estimation, Inference and Specification Analysis*. Econometric Society Monographs. 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.

A Proofs of Main Results

Proof of Proposition 3.1. We regard this proof as a special case of Proposition TA.1 in Appendix B of the Technical Appendix, by setting

$$\phi(x_t(\boldsymbol{\theta},\bar{x}),v_t,\boldsymbol{\theta}) = \omega + \alpha s_u(\hat{f}^u_t,u_t;\lambda) + \beta\,\hat{f}^u_t,$$

$v_t = u_t$, and $x_t(\boldsymbol{\theta},\bar{x}) = \hat{f}^u_t(\boldsymbol{\theta},\hat{f}^u_1)$. Here s_u is assumed to be $s_u \in \mathbb{C}^{(1,0,0)}(\mathcal{F}\times\mathcal{U}\times\Lambda)$ for convex $\mathcal{F}$, such that $\phi \in \mathbb{C}^{(1,0,0)}(\mathcal{X}\times\mathcal{V}\times\Theta)$ with a convex $\mathcal{X}$. Recall that $\{u_t\}_{t\in\mathbb{Z}}$ is an i.i.d. sequence by definition of the model. Conditions *(i)* and *(iii)* in Proposition TA.1 in Appendix B now directly follow from conditions *(i)* and *(iii)* of Proposition 3.1 (see the proof of Proposition 3.3 for a more thorough explanation). Condition *(iv)* in Proposition TA.1 directly follows from condition *(iv)* in Proposition 3.1 by observing that from the mean value theorem we have

$$\mathbb{E}r_1^k(\boldsymbol{\theta}) = \mathbb{E}\sup_{(x,x')\in\mathcal{X}\times\mathcal{X}:x\neq x'}\frac{|\phi(x,v_t,\boldsymbol{\theta})-\phi(x',v_t,\boldsymbol{\theta})|^k}{|x-x'|^k} \quad \leq$$
$$\mathbb{E}\sup_{x^*\in\mathcal{X}}\left|\frac{\partial\phi(x^*,v_t,\boldsymbol{\theta})}{\partial x}\right|^k = \mathbb{E}\sup_{f^{u*}\in\mathcal{F}}\left|\beta+\alpha\frac{\partial s_u(f^{u*},u_t,\boldsymbol{\theta})}{\partial f^u}\right|^k = \mathbb{E}\rho_1^k(\boldsymbol{\theta}) \;\; \forall\; k>0.$$

The same argumentation can be used to show that condition *(ii)* of Proposition TA.1 follows from condition *(ii)* of Proposition 3.1. ∎

Proof of Proposition 3.3. The results for the sequence $\{\hat{f}_t\}$ are obtained by application of Proposition TA.3 in Appendix B with $v_t = y_t$ and $x_t(\boldsymbol{\theta}, \bar{x}) = \hat{f}_t(\boldsymbol{\theta}, \hat{f}_1)$ and $\phi(x_t, v_t, \boldsymbol{\theta}) = \omega + \alpha s(\hat{f}_t, y_t; \lambda) + \beta \hat{f}_t$.

Step 1, SE for f_t: Condition *(i)* of Proposition TA.3 holds, because

$$
\begin{aligned}
\mathbb{E} \log^+ \sup_{\boldsymbol{\theta}\in\Theta} |\phi(\bar{x}, v_t, \boldsymbol{\theta}) - \bar{x}| &= \mathbb{E} \log^+ \sup_{\boldsymbol{\theta}\in\Theta} |\omega + \alpha s(\hat{f}_1, y_t; \lambda) + \beta \hat{f}_1 - \hat{f}_1| \\
&\leq \mathbb{E} \log^+ \sup_{\boldsymbol{\theta}\in\Theta} \Big[|\omega| + |\alpha| \cdot |s(\hat{f}_1, y_t; \lambda)| + |\beta - 1| \cdot |\hat{f}_1| \Big] \\
&\leq \log^+ \sup_{\omega\in\Omega} |\omega| + \log^+ \sup_{\alpha\in\mathcal{A}} |\alpha| + \mathbb{E} \log^+ \sup_{\lambda\in\Lambda} |s(\hat{f}_1, y_t; \lambda)| \\
&\qquad + \sup_{\beta\in\mathcal{B}} \log^+ |(\beta - 1)| + \log^+ |\hat{f}_1| < \infty
\end{aligned}
$$

with $\log^+ \sup_{\omega\in\Omega} |\omega| < \infty$, $\log^+ \sup_{\alpha\in\mathcal{A}} |\alpha| < \infty$ and $\sup_{\beta\in\mathcal{B}} \log^+ |(\beta - 1)| < \infty$ by compactness of Θ, and $\log^+ |\hat{f}_1| < \infty$ for any $\hat{f}_1 \in \mathcal{F} \subseteq \mathbb{R}$, and $\mathbb{E} \log^+ \sup_{\lambda\in\Lambda} |s(\hat{f}_1, y_t; \lambda)| < \infty$ by condition *(i)* in Proposition 3.3.
Condition *(ii)* in Proposition TA.3 holds, because

$$
\begin{aligned}
&\mathbb{E} \log \sup_{\boldsymbol{\theta}\in\Theta} r_1^1(\boldsymbol{\theta}) = \\
&\quad \mathbb{E} \log \sup_{\boldsymbol{\theta}\in\Theta} \sup_{(f,f')\in\mathcal{F}\times\mathcal{F}: f\neq f'} \frac{|\omega - \omega + \alpha(s(f, y_t; \lambda) - s(f', y_t; \lambda)) + \beta(f - f')|}{|f - f'|} \\
&\quad = \mathbb{E} \log \sup_{\boldsymbol{\theta}\in\Theta} \sup_{(f,f')\in\mathcal{F}\times\mathcal{F}: f\neq f'} \frac{|\alpha(s(f, y_t; \lambda) - s(f', y_t; \lambda)) + \beta(f - f')|}{|f - f'|} \\
&\quad = \mathbb{E} \log \sup_{\boldsymbol{\theta}\in\Theta} \sup_{(f,f')\in\mathcal{F}\times\mathcal{F}: f\neq f'} \frac{|\alpha \dot{s}_{y,t}(f^*; \lambda)(f - f') + \beta(f - f')|}{|f - f'|} \\
&\quad \leq \mathbb{E} \log \sup_{\boldsymbol{\theta}\in\Theta} \sup_{f^*\in\mathcal{F}} \Big| \alpha \dot{s}_{y,t}(f^*; \lambda) + \beta \Big| = \mathbb{E} \log \sup_{\boldsymbol{\theta}\in\Theta} \bar{\rho}_1(\boldsymbol{\theta}) < 0,
\end{aligned}
$$

where the second equality holds by the mean value theorem and where the last inequality follows directly from condition *(ii)* in Proposition 3.3.

Step 2, moment bounds for f_t: By a similar argument as in Step 1, we can show that condition *(iv)* in Proposition TA.3 follows from condition *(iv)* in Proposition 3.3. Condition *(iii)* in Proposition TA.3 for $n = n_f$ can be shown by noting that $\|\phi(\bar{x}, v_t, \cdot)\|_{n_f}^{\Theta} < \infty$ is implied by $(\|\phi(\bar{x}, v_t, \cdot)\|_{n_f}^{\Theta})^{n_f} < \infty$. The result now follows since by the C_r-inequality in Loève (1977, p.157), there exists a $0 < c < \infty$ such that

$$
\begin{aligned}
(\|\phi(\bar{x}, v_t, \cdot)\|_{n_f}^{\Theta})^{n_f} &= \mathbb{E} \sup_{\boldsymbol{\theta}\in\Theta} |\omega + \alpha s(\hat{f}_1, y_t; \lambda) + \beta \hat{f}_1|^{n_f} \\
&\leq c \cdot \sup_{\boldsymbol{\theta}\in\Theta} |\omega + \beta \hat{f}_1|^{n_f} + c \cdot |\alpha|^{n_f} \, \mathbb{E} \sup_{\boldsymbol{\theta}\in\Theta} |s(\hat{f}_1, y_t; \lambda)|^{n_f} < \infty,
\end{aligned}
$$

where the last inequality follows from condition *(iii)* in Proposition 3.3. ∎

Proof of Proposition 3.5. Step 1, SE for derivatives of f_t: The desired result follows by noting that the vector derivative processes $\{\hat{\boldsymbol{f}}_t^{(i)}\}_{t\in\mathbb{N}}$ for $i = 1, 2$ and initialized at $\hat{\boldsymbol{f}}_1^{(0:i)}$ satisfy the conditions of Theorem 2.10 in Straumann and Mikosch (2006) for perturbed stochastic recurrence equations, under the supremum norm $\|\cdot\|^{\Theta} = \sup_{\boldsymbol{\theta}\in\Theta}|\cdot|$. In particular, they consider a recurrence of the from $x_{t+1} = \phi_t(x_t)$ where $\{\phi_t\}$ converges to an SE sequence $\{\tilde{\phi}_t\}$ that satisfies the conditions of Bougerol's theorem $\mathbb{E}\log^+\sup_{\boldsymbol{\theta}\in\Theta}|\tilde{\phi}_t(0)| < \infty$, $\mathbb{E}\log\sup_{\boldsymbol{\theta}\in\Theta}\sup_x|\tilde{\phi}_t'(x)| < \infty$. In particular, one must have a logarithmic moment $\mathbb{E}\log^+\sup_{\boldsymbol{\theta}\in\Theta}|\tilde{x}_t|$ for the solution $\{\tilde{x}_t\}$ of the unperturbed SE system, and the perturbed recurrence must satisfy

$$\sup_{\boldsymbol{\theta}\in\Theta}|\phi_t(\bar{x}) - \tilde{\phi}_t(\bar{x})| \stackrel{e.a.s.}{\to} 0\ , \quad \text{for some } \bar{x}\in\mathbb{R} \quad \text{and} \quad \sup_{\boldsymbol{\theta}\in\Theta}\sup_x|\phi_t'(x) - \tilde{\phi}_t'(x)| \stackrel{e.a.s.}{\to} 0 \quad \text{as } t\to\infty.$$

Here we state the convergence of ϕ_t at some point $\bar{x}$ rather than at the origin $\phi_t(0)$ as in Straumann and Mikosch (2006) since our recursion (depending on the application) may not be well defined at $\bar{x} = 0$. As explained before, the perturbed sequence $\{\hat{\boldsymbol{f}}_t^{(i)}\}_{t\in\mathbb{N}}$ depends on the non-stationary sequences $\{\hat{f}_t\}_{t\in\mathbb{N}}$ and $\{\hat{\boldsymbol{f}}_t^{(1:i-1)}\}_{t\in\mathbb{N}}$, which are only stationary in the limit. The unperturbed initialized recurrence $\{\hat{\boldsymbol{f}}_t^{(i)}\}_{t\in\mathbb{N}}$ is equal in all respects, except that it instead depends on the limit SE filter $\{\boldsymbol{f}_t^{((0:i-1))}\}_{t\in\mathbb{Z}}$. The unperturbed limit process is denoted by $\{\boldsymbol{f}_t^{(i)}\}_{t\in\mathbb{Z}}$.

In Appendix D.2 we show that the dynamic equations generating each element of the partial derivative processes take the form

$$\hat{\boldsymbol{f}}_{t+1}^{(i)} \;=\; \boldsymbol{A}_t^{(i)}(\boldsymbol{\theta}, \hat{\boldsymbol{f}}_1^{(0:i-1)}) \;+\; \hat{\boldsymbol{f}}_t^{(i)}\; \boldsymbol{B}_t(\boldsymbol{\theta}, \hat{f}_1), \tag{A.1}$$

with $\boldsymbol{B}_t(\boldsymbol{\theta}, \hat{f}_1) = \beta + \alpha\,\partial s(\hat{f}_t(\boldsymbol{\theta}, \hat{f}_1), y_t; \lambda)/\partial f$ not depending on the order of the derivative i. The expressions for $\boldsymbol{A}_t^{(i)}(\boldsymbol{\theta}, \hat{f}_1)$ are presented in Appendix D.2 and only depend on derivatives up to order $\hat{\boldsymbol{f}}_t^{(i-1)}$. Note that $\boldsymbol{A}_t^{(i)}$ and $\boldsymbol{B}_t$ are written explicitly as a function of $\hat{\boldsymbol{f}}_1^{(0:i-1)}$ and $\hat{f}_1$ respectively, since they depend on the non-stationary filtered sequences $\{\hat{f}_t\}$ and $\{\hat{\boldsymbol{f}}_t^{(1:i-1)}\}$ initialized at $\hat{f}_1$ and $\hat{\boldsymbol{f}}_1^{(0:i-1)}$ respectively. In contrast, we let $\boldsymbol{A}_t^{(i)}(\boldsymbol{\theta})$ and $\boldsymbol{B}_t(\boldsymbol{\theta})$ denote the stationary counterparts of $\boldsymbol{A}_t^{(i)}(\boldsymbol{\theta}, \hat{\boldsymbol{f}}_1^{(0:i-1)})$ and $\boldsymbol{B}_t(\boldsymbol{\theta}, \hat{f}_1)$, respectively, that depend on the limit stationary filter $f_t(\boldsymbol{\theta})$. The recurrence convergence $\sup_{\boldsymbol{\theta}\in\Theta}|\phi_t(\bar{x}) - \tilde{\phi}_t(\bar{x})| \stackrel{e.a.s.}{\to} 0$ in Straumann and Mikosch (2006) corresponds here to having $\sup_{\boldsymbol{\theta}\in\Theta}|\boldsymbol{A}_t^{(i)}(\boldsymbol{\theta}, \hat{\boldsymbol{f}}_1^{(0:i-1)}) - \boldsymbol{A}_t^{(i)}(\boldsymbol{\theta})| \stackrel{e.a.s.}{\to} 0$ and $\sup_{\boldsymbol{\theta}\in\Theta}|\boldsymbol{B}_t(\boldsymbol{\theta}, \hat{f}_1) - \boldsymbol{B}_t(\boldsymbol{\theta})| \stackrel{e.a.s.}{\to} 0$. Both conditions are easily verified. We start by looking at the first derivative. Indeed, the expressions in Appendix D.2 show that $\boldsymbol{A}_t^{(1)}(\boldsymbol{\theta}, \hat{\boldsymbol{f}}_1^{(0:i-1)})$ satisfies

$$\sup_{\boldsymbol{\theta}\in\Theta}|\boldsymbol{A}_{j,t}^{(1)}(\boldsymbol{\theta}, \hat{f}_1) - \boldsymbol{A}_{j,t}^{(1)}(\boldsymbol{\theta})| \le \sup_f\sup_{\boldsymbol{\theta}\in\Theta}|\partial\boldsymbol{A}_{j,t}^{(1)}(\boldsymbol{\theta})/\partial f| \cdot \sup_{\boldsymbol{\theta}\in\Theta}|\hat{f}_t - f_t|$$

for each j and hence we obtain $\sup_{\boldsymbol{\theta}\in\Theta}|\boldsymbol{A}_{j,t}^{(1)}(\boldsymbol{\theta}, \hat{f}_1) - \boldsymbol{A}_{j,t}^{(1)}(\boldsymbol{\theta})| \stackrel{e.a.s.}{\to} 0$ by Lemma 2.1 in Straumann and Mikosch (2006) since $\sup_f\sup_{\boldsymbol{\theta}\in\Theta}|\partial\boldsymbol{A}_{j,t}^{(1)}(\boldsymbol{\theta})/\partial f|$ is SE with a logarithmic moment since $\min\{\bar{n}_s^f, \bar{n}_s^{\lambda f}\} > 0$ and because $\sup_{\boldsymbol{\theta}\in\Theta}|\hat{f}_t - f_t| \stackrel{e.a.s.}{\to} 0$ by Proposition 3.3. Similarly, we obtain

$$\sup_{\boldsymbol{\theta}\in\Theta}|\boldsymbol{B}_t(\boldsymbol{\theta}, \hat{f}_1) - \boldsymbol{B}_t(\boldsymbol{\theta})| \le \sup_f\sup_{\boldsymbol{\theta}\in\Theta}|\partial\boldsymbol{B}_t(\boldsymbol{\theta})/\partial f| \cdot \sup_{\boldsymbol{\theta}\in\Theta}|\hat{f}_t - f_t| \stackrel{e.a.s.}{\to} 0 \quad \text{as } t\to\infty,$$

since $\bar{n}_s^{ff} > 0$ implies that $\sup_f \sup_{\boldsymbol{\theta}\in\Theta} |\partial \boldsymbol{B}_t(\boldsymbol{\theta})/\partial f|$ is SE with with a logarithmic moment, and $\sup_{\boldsymbol{\theta}\in\Theta} |\hat{f}_t - f_t|$ vanishes e.a.s. The convergence of the Lipschitz coefficients $\sup_{\boldsymbol{\theta}\in\Theta} \sup_x |\phi_t'(x) - \tilde{\phi}_t'(x)| = |\boldsymbol{B}_t(\boldsymbol{\theta}, \hat{f}_1) - \boldsymbol{B}_t(\boldsymbol{\theta})| \overset{e.a.s.}{\to} 0$ follows immediately.

For the second derivative process, the same argument using Lemma 2.1 in Straumann and Mikosch (2006) applies sequentially. As the argument is slightly more subtle, we prove it in Lemma TA.17 of the Technical Appendix.

Finally, we note that the unperturbed recursions satisfy the conditions of Bougerol's theorem, which is implied by the verification of conditions (iii)-(iv) of Proposition TA.3 for the unperturbed system in the next step of the proof. The logarithmic moment of the SE limit process that we need also follows directly from the verification of these conditions, because it implies the existence of a moment of some positive order under the current conditions. In the notation of Straumann and Mikosch (2006), this means that the limit recursion $\tilde{\phi}_t$ is SE and that its solution $\{\tilde{x}_t\}_{t\in\mathbb{Z}}$ has a logarithmic moment uniformly over the parameter space. Thus, after we proved that these conditions hold, the e.a.s. convergence of the initialized perturbed sequence to an SE limit sequence uniformly over the parameter space follows from Theorem 2.10 in Straumann and Mikosch (2006).

Step 2, moment bounds for derivatives of f_t: To establish the existence of moments for the derivative processes, we need to verify that conditions *(iii)–(iv)* of Proposition TA.3 hold. For the limit derivative processes, we can apply Proposition TA.3 directly to the unperturbed system.

Inspection of the formula for $\boldsymbol{A}_{j,t}^{(1)}(\boldsymbol{\theta})$ reveals that $\boldsymbol{A}_{j,t}^{(1)}(\boldsymbol{\theta})$ has $n_{f_{\boldsymbol{\theta}}} = \min\{n_f\ ,\ n_s\ ,\ n_s^{\lambda}\}$ bounded moments and $\boldsymbol{A}_{j,t}^{(2)}(\boldsymbol{\theta})$ has $n_{f_{\boldsymbol{\theta}\boldsymbol{\theta}}}$ moments as defined in Proposition 3.5, which follows from Hölder's inequality. Inspection of the expression for $\boldsymbol{B}_t(\boldsymbol{\theta})$ and condition (iv) of Proposition 3.3 reveals that $\boldsymbol{B}_t(\boldsymbol{\theta})$ has n_f moments.

Thus, under the conditions of Proposition 3.5, condition *(iii)* in Proposition TA.3 holds with $n_{f_{\boldsymbol{\theta}}}$ moments for the first derivative process and $n_{f_{\boldsymbol{\theta}\boldsymbol{\theta}}}$ moments for the second derivative process, since for any $n > 0$, by the C_r-inequality in Loève (1977, p.157), there exists a $0 < c < \infty$ such that,

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

for each j and where for example $\boldsymbol{A}_{j,t}^{(i)}(\boldsymbol{\theta})$ denotes the j-th element of the vector or matrix $\boldsymbol{A}_t^{(i)}(\boldsymbol{\theta})$. Condition *(iv)* in Proposition TA.3 holds for $i = 1, 2$, since for any pair $(f_j^{(i)}, f_j^{(i)\prime}) \in \mathcal{F} \times \mathcal{F} : f_j^{(i)} \neq$

$f_j^{(i)\prime}$:

$$\begin{aligned}\sup_{\boldsymbol{\theta}\in\Theta}\frac{\left|\phi(f_j^{(i)},v_t,\boldsymbol{\theta})-\phi(f_j^{(i)\prime},v_t,\boldsymbol{\theta})\right|}{\left|f_j^{(i)}-f_j^{(i)\prime}\right|}&=\sup_{\boldsymbol{\theta}\in\Theta}\frac{\left|\boldsymbol{B}_t(\boldsymbol{\theta})(f_j^{(i)}-f_j^{(i)\prime})\right|}{\left|f_j^{(i)}-f_j^{(i)\prime}\right|}\\&=\sup_{\boldsymbol{\theta}\in\Theta}\left|\beta+\alpha\frac{s(f_t,y_t;\lambda)}{\partial f}\right|\\&\leq\sup_{(f^*,y,\boldsymbol{\theta})\in\mathcal{F}\times\mathcal{Y}\times\Theta}\left|\ \beta+\alpha\,\frac{\partial s(f^*,y;\lambda)}{\partial f}\ \right|<1\,,\end{aligned}$$

for every j and where $v_t=(\boldsymbol{f}_t^{(0:i-1)},y_t)$. The first equality holds because we are working with the unperturbed sequence and the final inequality holds because of condition *(iv)* of Proposition 3.3. We thus obtain, by Proposition TA.3, $n_{f_{\boldsymbol{\theta}}}$ ($n_{f_{\boldsymbol{\theta}\boldsymbol{\theta}}}$) moments for the first (second) derivative limit process. ∎

Proof of Theorem 4.3. The result follows immediately from the differentiability of $\bar{p}$, $\bar{g}$, $\bar{g}'$, the compactness of Θ, and the Weierstrass theorem. For a detailed proof, see Technical Appendix B. ∎

Proof of Theorem 4.6. Following the classical consistency argument found in for instance White (1994, Theorem 3.4) or Gallant and White (1988, Theorem 3.3), we obtain $\hat{\boldsymbol{\theta}}_T(\hat{f}_1)\overset{a.s.}{\to}\boldsymbol{\theta}_0$ from the uniform convergence of the criterion function and the identifiable uniqueness of the maximizer $\boldsymbol{\theta}_0\in\Theta$,

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

Step 1, uniform convergence: Let $\ell_T(\boldsymbol{\theta})$ denote the likelihood $\ell_T(\boldsymbol{\theta},\hat{f}_1)$ with $\hat{f}_t$ replaced by f_t. 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 t-th observation to the likelihood function ℓ_T. We have

$$\begin{aligned}&\sup\nolimits_{\boldsymbol{\theta}\in\Theta}|\ell_T(\boldsymbol{\theta},\hat{f}_1)-\ell_\infty(\boldsymbol{\theta})|\ \leq\\&\sup\nolimits_{\boldsymbol{\theta}\in\Theta}|\ell_T(\boldsymbol{\theta},\hat{f}_1)-\ell_T(\boldsymbol{\theta})|+\sup\nolimits_{\boldsymbol{\theta}\in\Theta}|\ell_T(\boldsymbol{\theta})-\ell_\infty(\boldsymbol{\theta})|.\end{aligned}\qquad\text{(A.2)}$$

The first term vanishes by application of Lemma 2.1 in Straumann and Mikosch (2006) since $\hat{f}_t$ converges e.a.s. to f_t and $\sup_{\boldsymbol{\theta}\in\Theta}\sup_f|\nabla\ell_T(\boldsymbol{\theta})|$ has a logarithmic moment because $\bar{n}_\nabla>0$. The second term vanishes by Rao (1962); see Lemmas TA.5 and TA.6 form Technical Appendix B, respectively.

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

Proof of Theorem 4.10. We index the true $\{f_t\}$ and the observed random sequence $\{y_t\}$ by the parameter $\boldsymbol{\theta}_0$, e.g. $\{y_t(\boldsymbol{\theta}_0)\}$, since under correct specification the observed data is a subset of the realized path of a stochastic process $\{y_t\}_{t\in\mathbb{Z}}$ generated by a score-driven model under $\boldsymbol{\theta}_0 \in \Theta$. As conditions *(i)* and *(ii)* of Proposition 3.1 hold immediately by Assumption 4.9 and condition *(v)* follows immediately from the i.i.d. exogenous nature of the sequence $\{u_t\}$, it follows by Proposition 3.1 that the true sequence $\{f_t(\boldsymbol{\theta}_0)\}$ is SE and has at least n_f moments for any $\boldsymbol{\theta} \in \Theta$. The SE nature and n_f moments of $\{f_t(\boldsymbol{\theta}_0)\}$ together with part *(iii)* of Assumption 4.8 imply, in turn, that $\{y_t(\boldsymbol{\theta}_0)\}$ is SE with $n_y = n_g$ moments.

Step 1 (formulation and existence of the limit criterion $Q_\infty(\boldsymbol{\theta})$): As shown in the proof of Theorem 4.6, the limit criterion function $Q_\infty(\boldsymbol{\theta})$ is well-defined for every $\boldsymbol{\theta} \in \Theta$ by

$$Q_\infty(\boldsymbol{\theta}) = \mathbb{E}\tilde{\ell}_t(\boldsymbol{\theta}) = \mathbb{E}\log p_{y_t|y_{t-1},y_{t-2},\ldots}\Big(y_t(\boldsymbol{\theta}_0)\Big|y_{t-1}(\boldsymbol{\theta}_0), y_{t-2}(\boldsymbol{\theta}_0), \ldots; \boldsymbol{\theta}\Big).$$

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

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

To do this, we use Lemma TA.7 from Technical Appendix B and rewrite

$$\begin{aligned} Q_\infty(\boldsymbol{\theta}) - Q_\infty(\boldsymbol{\theta}_0) = \\ \int\!\!\int \left[\int p_y(y|f,\lambda_0)\log\frac{p_y(y|\tilde{f};\lambda)}{p_y(y|f;\lambda_0)}\,\mathrm{d}y\right] p_{f_t,\tilde{f}_t}(f,\tilde{f};\boldsymbol{\theta}_0,\boldsymbol{\theta})\,\mathrm{d}f\,\mathrm{d}\tilde{f}, \end{aligned} \tag{A.3}$$

for all $(\boldsymbol{\theta}_0, \boldsymbol{\theta}) \in \Theta\times\Theta : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, where $p_{f_t,\tilde{f}_t}(f,\tilde{f};\boldsymbol{\theta}_0,\boldsymbol{\theta})$ is the bivariate pdf for the pair $(f_t(\boldsymbol{\theta}_0), \tilde{f}_t(\boldsymbol{\theta}))$. We note that the pdf $p_{f_t,\tilde{f}_t}(f,\tilde{f};\boldsymbol{\theta}_0,\boldsymbol{\theta})$ depends on both $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}$, as for instance the recursion defining $\tilde{f}_t(\boldsymbol{\theta})$ depends on both $\boldsymbol{\theta}$ and on $y_t(\boldsymbol{\theta}_0)$, which in turn depends on $\boldsymbol{\theta}_0$. Next, we use Gibb's inequality to show that this quantity is negative for $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$.

Step 2 (use of Gibb's inequality): Gibb's inequality ensures that, for any given $(f, \tilde{f}, \lambda_0, \lambda) \in \mathcal{F} \times \tilde{\mathcal{F}} \times \Lambda \times \Lambda$, the inner integral in (A.3) satisfies

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

with equality holding if and only if $p_y(y|\tilde{f};\lambda) = p_y(y|f;\lambda_0)$ almost everywhere in $\mathcal{Y}$ with respect to $p_y(y|f,\lambda_0)$. By Lemma TA.8 from Technical Appendix B there exists a set $\mathit{YF\tilde{F}} \subseteq \mathcal{Y} \times \mathcal{F} \times \tilde{\mathcal{F}}$ with positive probability mass and with orthogonal projections $\mathit{Y\tilde{F}} \subseteq \mathcal{Y} \times \mathcal{F}$, $\mathit{F\tilde{F}} \subseteq \mathcal{F} \times \tilde{\mathcal{F}}$, etc., for which (i)–(ii) hold if $\lambda \neq \lambda_0$, and for which (i)–(iii) hold if $\lambda = \lambda_0$, where

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

(ii) if $(\tilde{f}, \lambda) \neq (f, \lambda_0)$, then $p_y(y|\tilde{f};\lambda) \neq p_y(y|f;\lambda_0)\ \forall\ (y, f, \tilde{f}) \in \mathit{YF\tilde{F}}$;

(iii) if $\lambda = \lambda_0$ and $(\omega, \alpha, \beta) \neq (\omega_0, \alpha_0, \beta_0)$, then $f \neq \tilde{f}$ for every $(f, \tilde{f}) \in \mathit{F\tilde{F}}$.

Hence, if $\lambda \neq \lambda_0$, the strict Gibb's inequality follows directly from (i) and (ii) and the inner integral and the fact that $YF\tilde{F}$ has positive probability mass. If $\lambda = \lambda_0$, property (iii) ensures $f \neq \tilde{f}$ on a subset $F\tilde{F}$ with positive probability mass, and hence the strict inequality again follows via (ii) and (i). ■

Proof of Corollary 4.11. The desired result is obtained by showing (i) that under the maintained assumptions, $\{y_t\}_{t\in\mathbb{Z}} \equiv \{y_t(\boldsymbol{\theta}_0)\}_{t\in\mathbb{Z}}$ is an SE sequence satisfying $\mathbb{E}|y_t(\boldsymbol{\theta}_0)|^{n_y} < \infty$; (ii) that $\boldsymbol{\theta}_0 \in \Theta$ is the unique maximizer of $\ell_\infty(\boldsymbol{\theta}, \hat{f}_1)$ on Θ; and then (iii) appealing to Theorem 4.6. The fact that $\{y_t(\boldsymbol{\theta}_0)\}_{t\in\mathbb{Z}}$ is an SE sequence is obtained by applying Proposition 3.1 under Assumptions 4.8 and 4.9 to ensure that $\{\hat{f}_t(\boldsymbol{\theta}_0)\}_{t\in\mathbb{N}}$ converges e.a.s. to an SE limit $\{f_t(\boldsymbol{\theta}_0)\}_{t\in\mathbb{Z}}$ satisfying $\mathbb{E}|f_t(\boldsymbol{\theta}_0)|^{n_{f^u}} < \infty$. This implies by continuity of g on $\mathcal{F}\times\mathcal{U}$ (implied by $\bar{g} \in \mathbb{C}^{(2,0)}(\bar{\mathcal{F}}\times\mathcal{Y})$ in Assumption 4.2) that $\{y_t(\boldsymbol{\theta}_0)\}_{t\in\mathbb{Z}}$ is SE. Furthermore, Assumption 4.8 implies that $\mathbb{E}|y_t(\boldsymbol{\theta}_0)|^{n_y} < \infty$ for $n_y = n_g$. Finally, the uniqueness of $\boldsymbol{\theta}_0$ is obtained by applying Theorem 4.10 under Assumptions 4.8 and 4.9. ■

Proof of Theorem 4.15. 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 $\ell_T(\boldsymbol{\theta}, \hat{f}_1)$ in $\boldsymbol{\theta} \in \Theta$;
(iii) the asymptotic normality of the score

$$\sqrt{T}\ell'_T\big(\boldsymbol{\theta}_0, \hat{\boldsymbol{f}}_1^{(0:1)}\big) \xrightarrow{d} \text{N}(0, \mathcal{J}(\boldsymbol{\theta}_0)), \qquad \mathcal{J}(\boldsymbol{\theta}_0) = \lim_{T\to\infty} T^{-1}\mathbb{E}\Big(\sum_{t=1}^{T} \tilde{\ell}'_t(\boldsymbol{\theta}_0)\Big)\Big(\sum_{t=1}^{T} \tilde{\ell}'_t(\boldsymbol{\theta}_0)^\top\Big); \tag{A.4}$$

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

$$\sup_{\boldsymbol{\theta}\in\Theta} \big\|\ell''_T(\boldsymbol{\theta}, \hat{\boldsymbol{f}}_1^{(0:2)}) - \ell''_\infty(\boldsymbol{\theta})\big\| \xrightarrow{a.s.} 0; \tag{A.5}$$

(v) the non-singularity of the limit $\ell''_\infty(\boldsymbol{\theta}_0) = \mathbb{E}\tilde{\ell}''_t(\boldsymbol{\theta}_0) = \mathcal{I}(\boldsymbol{\theta}_0)$, which holds by assumption.

Step 1 (consistency and differentiability): Consistency to an internal point of Θ follows immediately by Theorem 4.6 and the additional assumption that $\boldsymbol{\theta}_0 \in \text{int}(\Theta)$. The differentiability of the likelihood function follows directly by Assumption 4.2 and the expressions for the likelihood in Technical Appendix D.

Step 2, CLT: The asymptotic normality of the score $\ell'_T\big(\boldsymbol{\theta}_0, \hat{\boldsymbol{f}}_1^{(0:1)}\big)$ in (A.4) follows by applying a CLT to $\ell'_T(\boldsymbol{\theta}_0)$,

$$\sqrt{T}\ell'_T(\boldsymbol{\theta}_0) \xrightarrow{d} \text{N}(0, \mathcal{J}(\boldsymbol{\theta}_0)), \qquad \mathcal{J}(\boldsymbol{\theta}_0) = \lim_{T\to\infty} T^{-1}\mathbb{E}\Big(\sum_{t=1}^{T} \tilde{\ell}'_t(\boldsymbol{\theta}_0)\Big)\Big(\sum_{t=1}^{T} \tilde{\ell}'_t(\boldsymbol{\theta}_0)^\top\Big) < \infty, \tag{A.6}$$

and by showing that the effect of initial conditions vanishes, i.e.,

$$\sqrt{T}\|\ell_T'(\boldsymbol{\theta}_0, \hat{\boldsymbol{f}}_1^{(0:1)}) - \ell_T'(\boldsymbol{\theta}_0)\| \xrightarrow{a.s.} 0 \quad \text{as} \quad T \to \infty. \tag{A.7}$$

and by appealing to Theorem 18.10[iv] in van der Vaart (2000). We note that the CLT for SE martingale difference sequences (mds) in Billingsley (1961) cannot be used to obtain (A.6) as we allow for model misspecification, and hence the mds property need not hold. Instead, we obtain (A.6) by applying the CLT for SE NED sequences in Davidson (1992, 1993) (see also Davidson, 1994; Pötscher and Prucha, 1997). Lemma TA.11 in Technical Appendix F ensures that the score $\ell_T'(\boldsymbol{\theta}_0)$ is a sample average of a sequence that is SE and NED of size -1 on a strongly mixing sequence. In addition, the existence of $\mathcal{J}(\boldsymbol{\theta}_0)$ follows from Lemma TA.9 and the assumption that $n_{\ell'} > 2$ in Assumption 4.12. Finally, the a.s. convergence in (A.7) follows directly by Lemma TA.12 in Technical Appendix F.

Step 3, uniform convergence of ℓ'': The proof of the uniform convergence in (iv) is similar to that of Theorem 4.6. We have

$$\sup_{\boldsymbol{\theta}\in\Theta} \|\ell_T''(\boldsymbol{\theta}, \hat{\boldsymbol{f}}_1^{(0:2)}) - \ell_\infty''(\boldsymbol{\theta})\| \leq \sup_{\boldsymbol{\theta}\in\Theta} \|\ell_T''(\boldsymbol{\theta}, \hat{\boldsymbol{f}}_1^{(0:2)}) - \ell_T''(\boldsymbol{\theta})\| + \sup_{\boldsymbol{\theta}\in\Theta} \|\ell_T''(\boldsymbol{\theta}) - \ell_\infty''(\boldsymbol{\theta})\|. \tag{A.8}$$

The first term on the right-hand side of (A.8) vanishes a.s. which can be shown by considering the separate terms in the expression of ℓ_T''; see Lemma TA.13 in Technical Appendix F.

For the second term in (A.8) we use the same approach as Lemma TA.6, meaning that we apply the ergodic theorem for separable Banach spaces of Rao (1962) to $\{\tilde{\ell}_t(\cdot)\}$. So the term converges if $\{\ell_T''\}_{t\in\mathbb{Z}}$ is and SE and $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\|\tilde{\ell}_t''(\boldsymbol{\theta})\| < \infty$. The former is implied by continuity of ℓ'' on the SE sequence $\{(y_t, \boldsymbol{f}_t^{(0:2)}(\cdot))\}_{t\in\mathbb{Z}}$ and Proposition 4.3 in Krengel (1985), where $\{(y_t, \boldsymbol{f}_t^{(0:2)}(\cdot))\}_{t\in\mathbb{Z}}$ is SE by Proposition 3.3 under the maintained assumptions. The moment bound $\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}\|\tilde{\ell}_t''(\boldsymbol{\theta})\| < \infty$ follows from $n_{\ell''} \geq 1$ in Assumption 4.12 and Lemma TA.10 in Technical Appendix F.

∎

Proof of Theorem 4.16. The desired result is obtained by applying Corollary 4.11 to guarantee that under the maintained assumptions $\{y_t\}_{t\in\mathbb{Z}} \equiv \{y_t(\boldsymbol{\theta}_0)\}_{t\in\mathbb{Z}}$ is an SE sequence satisfying $\mathbb{E}|y_t(\boldsymbol{\theta}_0)|^{n_y} < \infty$ for $n_y \geq 0$, and that $\boldsymbol{\theta}_0 \in \Theta$ is the unique maximizer of $\ell_\infty(\boldsymbol{\theta}, \hat{f}_1)$ on Θ. Then the statement follows along the same lines as the proof of Theorem 4.15. Note that the non-singularity of the limit $\ell_\infty''(\boldsymbol{\theta}_0) = \mathbb{E}\tilde{\ell}_t''(\boldsymbol{\theta}_0) = \mathcal{I}(\boldsymbol{\theta}_0)$ is implied by Theorem 1 in Rothenberg (1971), because the model is correctly specificied, $\boldsymbol{\theta}_0$ is the unique maximizer of $\ell_\infty(\boldsymbol{\theta})$ in Θ and the assumption that $\mathbb{E}\tilde{\ell}_t''(\boldsymbol{\theta}_0)$ is regular in the sense of Rothenberg (1971). ∎