

TI 2014-029/III  
Tinbergen Institute Discussion Paper



# Maximum Likelihood Estimation for Score-Driven Models

**REVISION: 23 October 2017**

*Francisco Blasques<sup>1</sup>*

*Siem Jan Koopman<sup>1,2</sup>*

*André Lucas<sup>1</sup>*

<sup>1</sup>Vrije Universiteit Amsterdam School of Business and Economics; Tinbergen Institute  
Amsterdam, The Netherlands

<sup>2</sup>CREATES, Aarhus University

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

Contact: [discussionpapers@tinbergen.nl](mailto:discussionpapers@tinbergen.nl)

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

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam  
Gustav Mahlerplein 117  
1082 MS Amsterdam  
The Netherlands  
Tel.: +31(0)20 598 4580

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

# Maximum Likelihood Estimation for Score-Driven Models\*

Francisco Blasques<sup>(a)</sup>, Siem Jan Koopman<sup>(a,b)</sup>, André Lucas<sup>(a)</sup>

<sup>(a)</sup> Vrije Universiteit Amsterdam School of Business and Economics,  
Tinbergen Institute Amsterdam

<sup>(b)</sup> CREATES, Aarhus University

23 October 2017

## Abstract

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

*Some key words:* score-driven models, time-varying parameters, Markov processes, stationarity, invertibility, consistency, asymptotic normality.

## 1 Introduction

The class of score-driven models encompasses many well-known time-varying parameter models in the literature, such as 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. Furthermore, the score-driven models have given rise to a new strand of empirical literature, which includes dynamic models for location and scale

---

\*We thank Peter Boswijk, Christian Francq, Andrew Harvey, and Anders Rahbek, as well as the 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, and University of Maastricht, for helpful comments and discussions. Blasques and Lucas thank the Dutch National Science Foundation (NWO; grant VICI453-09-005) for financial support. Koopman acknowledges support from CREATES, 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>.

of fat-tailed data (Harvey and Luati, 2014), models for mixed measurement dynamic factor structures (Creal et al., 2014), models for dynamic spatial processes (Blasques et al., 2016; Catania and Billé, 2017), and models for 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, 2017), and with realized measures as inputs (De Lira Salvatierra and Patton, 2015; Opschoor et al., 2017). Despite the impressive wide range of new models with score-driven dynamics, not many theoretical results are available for the estimators of the parameters in such models. The main complication lies in the non-linearity of the updating equation in score-driven models. In this paper, we attempt to fill this gap by deriving new asymptotic results for the maximum likelihood estimator which are applicable to a wide class of score-driven models.

In general, a score-driven model is specified by the equations

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

where  $y_t$  denotes the observed data,  $f_t$  is a time-varying parameter characterizing the conditional density  $p_y$  of  $y_t$ ,  $\nabla_t := \nabla(f_t, y_t; \lambda)$  denotes score of the predictive conditional density, and  $s_t$  denotes the scaled score for some choice of scaling function  $S(f_t; \lambda)$ . The static parameters in the density  $p_y$  are collected in a vector  $\lambda$  while all static parameters, including the updating coefficients  $\omega$ ,  $\alpha$ , and  $\beta$ , are gathered in the parameter vector  $\boldsymbol{\theta} = (\omega, \alpha, \beta, \lambda^\top)^\top$  with  $^\top$  denoting transposition.

A distinguishing feature of the model in (1.1) is the use of the scaled score  $s_t$  in the transition equation for  $f_t$ . A popular example of a score-driven model is the Student's  $t$  conditional volatility model, given by

$$\begin{aligned} y_t &= f_t^{1/2} \cdot u_t, \\ f_{t+1} &= \omega + \alpha (w_t y_t^2 - f_t) + \beta f_t, \\ w_t &= (1 + \lambda^{-1}) / (1 + \lambda^{-1} y_t^2 / f_t), \end{aligned} \tag{1.2}$$

where  $\{u_t\}_{t \in \mathbb{Z}}$  is a sequence of independent identically distributed random variables with Student's  $t$  distribution and  $\lambda$  degrees of freedom; see Creal et al. (2011, 2013) and Harvey (2013). The time-varying parameter  $f_t$  represents the conditional volatility of  $y_t$ . We use this model as our leading example throughout the paper.

The score model in (1.2) is markedly different from a GARCH model with Student's  $t$  innovations. In particular, next period's volatility is not merely driven by the square  $y_t^2$  of past observations, but rather by the weighted squares using weights  $(1 + \lambda^{-1}) / (1 + \lambda^{-1} y_t^2 / f_t)$ . These weights tend to zero for large values of  $y_t^2$ . The score-driven dynamics thus automatically correct for outliers if  $y_t$  is allowed to be conditionally fat-tailed. This is a desirable feature for many financial time series. In the limit as  $\lambda \rightarrow \infty$ , the weight  $w_t$  collapses to unity, such that we recover the GARCH model. More details on our leading example and other examples are presented in Section 2.1.

The score-driven model in equation (1.1) is observation-driven in the classification of Cox (1981). Therefore, maximum likelihood estimation of the static parameter vector can easily

be achieved via a prediction error decomposition. In particular, the likelihood function is known 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, using the model confidence set approach of Hansen et al. (2011), 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 data generating process.

We establish an asymptotic statistical theory for the maximum likelihood estimator of the static parameters of score-driven models as specified in (1.1). Our results have a number of distinctive features compared to earlier theoretical contributions on observation-driven models and, more specifically, on score-driven models.

First, the asymptotic properties that we derive for the maximum likelihood estimator (MLE) are global in nature. For example, we provide a global identification result for score-driven models in terms of low-level conditions. In particular, we ensure that the likelihood function has a unique global maximum over the entire parameter space. Our global result differs from the existing literature that 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) and Harvey (2013). Second, our effort to obtain primitive *low-level* conditions that are formulated in terms of the basic structure of the model extends well beyond the global identification conditions mentioned above. For instance, we obtain the required moments of the likelihood function directly from assumptions concerning the properties of the basic building blocks of equation (1.1), such as the shape of the density function  $p_y$ . The use of primitive conditions is typically useful for empirical researchers who want to establish the asymptotic properties of the MLE for their own score model. We are able to obtain low-level conditions by adapting Theorem 3.1 in Bougerol (1993). The adapted theorem not only delivers the strict stationarity and ergodicity of stochastic sequences, but also produces bounded moments of any desired order 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. These results allow us to establish the invertibility of the score filter and obtain our 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 mis-specified models. These results extend the literature for score-driven models, which focus only on the correctly specified case. By allowing for model mis-specification, we ‘align’ the asymptotic estimation theory for score-driven models with the existing information-theoretic optimality results of Blasques et al. (2015).

The newly developed theory presented in this paper allows us to establish results for a much wider range of score-driven models than the ones currently studied, which typically focus on fat-tailed models with log likelihoods that have uniformly bounded third order derivatives; see e.g. 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 all the existing results on score-driven models, which 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. (2016), among others. Without invertibility, all the existing asymptotic results on score-driven models must implicitly assume that the first value of the true time-varying parameter,  $f_1$ , is random *and* known exactly, while the remaining sequence  $\{f_t\}_{t \geq 2}$  is unobserved. This seems unrealistic.

The lack of theoretical results for the class of score-driven models defined in (1.1) stands in sharp contrast to the large number of results available for the MLE in GARCH models. We do not attempt to review that literature here; for good overviews, see for instance Straumann (2005) or Francq and Zakoian (2010). The main cause for the limited theoretical progress for score-driven models lies in their typical 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 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 an intimate link between the two that departs from most of the literature, where the properties of the likelihood function and the properties of the time-varying parameter dynamics can be dealt with separately.

The remainder of this paper is organized as follows. Section 2 introduces the model. In Section 3, we obtain stationarity, ergodicity, invertibility, and existence of moments of filtered score-driven sequences using primitive conditions. Section 4 proves global identification, consistency, and asymptotic normality of the MLE. Section 5 concludes. The proofs of the main theorems are gathered in the Appendix. More technical background material is relegated to the online Technical Appendix (TA).

## 2 The Score-Driven Model

### 2.1 An Introduction

As an intuitive introductory example of score-driven models, we consider a score-driven time-varying location model of the form

$$y_t = f_t + u_t, \quad u_t \sim p_u(u_t; \lambda), \quad (2.1)$$

$$f_{t+1} = \omega + \alpha \nabla_t + \beta f_t, \quad \nabla_t = \frac{\partial \log p_u(y_t - f_t; \lambda)}{\partial f_t} \quad (2.2)$$

where  $f_t$  is the time-varying location parameter,  $\log$  denotes the natural logarithm, and  $p_u$  is the error density that depends on an unknown, static parameter vector  $\lambda$ . For simplicity, we have set the scaling function from equation (1.1) in this example to  $S(f_t; \lambda) = 1$ . The dynamics of  $f_{t+1}$  can easily be extended by including higher order lags of  $f_t$  and/or of the score  $\nabla_t = \partial \log p_u(y_t - f_t; \lambda) / \partial f_t$ , or by imposing structural time series dynamics or including exogenous variables; see for further details Creal et al. (2013), Harvey and Luati (2014), and Ord et al. (1997).

If  $p_u$  in equation (2.1) is the normal distribution with mean zero and variance  $\lambda$ , the score in equation (2.2) becomes  $(y_t - f_t) / \lambda$ ,

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

The location  $f_{t+1}$  then moves linearly up and down with the previous residual value  $y_t - f_t$  in an intuitive way. In fact, model (2.1)–(2.2) can be re-written as a standard linear autoregressive moving average (ARMA) model for  $y_t$  of order (1, 1) with autoregressive coefficient  $\beta$  and moving average coefficient  $(\alpha / \lambda) - \beta$ .

The score-driven model becomes more interesting if we consider a non-normal distribution, such as a Student's  $t$  distribution; see Creal et al. (2013) and Harvey and Luati (2014). In this case, the score-driven model gives rise to an outlier robust filtering approach. Consider a Student's  $t$  distribution with mean zero, squared scale parameter  $\lambda_1$  and degrees of freedom parameter  $\lambda_2$ . Equation (2.2) then reduces to

$$f_{t+1} = \omega + \alpha (\lambda_2 + 1) \frac{(y_t - f_t)}{\lambda_1 \lambda_2 + (y_t - f_t)^2} + \beta f_t, \quad (2.3)$$

such that the location parameter  $f_{t+1}$  now reacts nonlinearly to past values of  $f_t$  for fixed data  $y_t$ . For example, whereas a large residual  $(y_t - f_t)$  pushes the next location parameter  $f_{t+1}$  substantially up if the normal distribution is assumed ( $\lambda_2 \rightarrow \infty$ ), under the Student's  $t$  assumption the impact of such a large residual is of order  $(y_t - f_t)^{-1}$ . The impact on  $f_{t+1}$  therefore becomes negligible as  $(y_t - f_t)$  diverges. The intuition is that such observations are more likely due to the fat-tailedness property of the Student's  $t$  distribution while they are not very informative about gradual shifts in the location parameter.

To elucidate on how the theory as developed in this paper may be applied to a realistic case, we adopt the leading example of a score-driven time-varying volatility model of Creal et al. (2011, 2013). By providing the details for this example, we keep the exposition focused. The application of the theory is, however, not limited to this particular case. A substantial range of additional illustrations of the theory can be provided, including robust time-varying location models with skewed distributions; conditional volatility models for log volatility; extensions of conditional duration (ACD) models of Engle and Russell (1998), Grammig and Maurer (2000), Bauwens and Giot (2000), and Koopman et al. (2016) that allow for fat-tailed densities; and time-varying location models with non-linear transformations of the location parameter, thus extending examples studied by Harvey (2013) and Harvey and Luati (2014).

**Main example:.** *Consider the conditional Student's  $t$  model with time varying volatility as given by*

$$\begin{aligned} y_t &= f_t^{1/2} \cdot u_t, \\ f_{t+1} &= \omega + \alpha \left( \left[ (1 + \lambda^{-1}) / (1 + \lambda^{-1} y_t^2 / f_t) \right] y_t^2 - f_t \right) + \beta f_t. \end{aligned} \quad (2.4)$$

where  $\{u_t\}_{t \in \mathbb{Z}}$  is a sequence of independent identically distributed random variables from the Student's  $t$  distribution with  $\lambda$  degrees of freedom. We ensure positivity of the scale  $f_t$  by imposing  $\beta > \alpha > 0$ ,  $\omega > 0$ , and  $\bar{f}_1 > 0$ , where  $\bar{f}_1$  is the initial condition for  $f_t$  at time  $t = 1$ . The raw score is scaled by a factor proportional to the inverse conditional Fisher information as suggested by Creal et al. (2013), to account for the local curvature of predictive density at time  $t$ , i.e.,  $S(f_t; \lambda) = 2f_t^2$ . For the Gaussian case ( $\lambda \rightarrow \infty$ ) we recover a slightly reparameterized version of the standard GARCH model,  $f_{t+1} = \omega + \alpha y_t^2 + (\beta - \alpha)f_t$ .

Figure 1 provides a typical illustration of the difference between the GARCH dynamics and the score-driven dynamics in (2.4). Large spikes in  $y_t^2$ , particularly incidental ones such as those at the end of 1987 and 1989, result in a spike in the GARCH volatility, followed

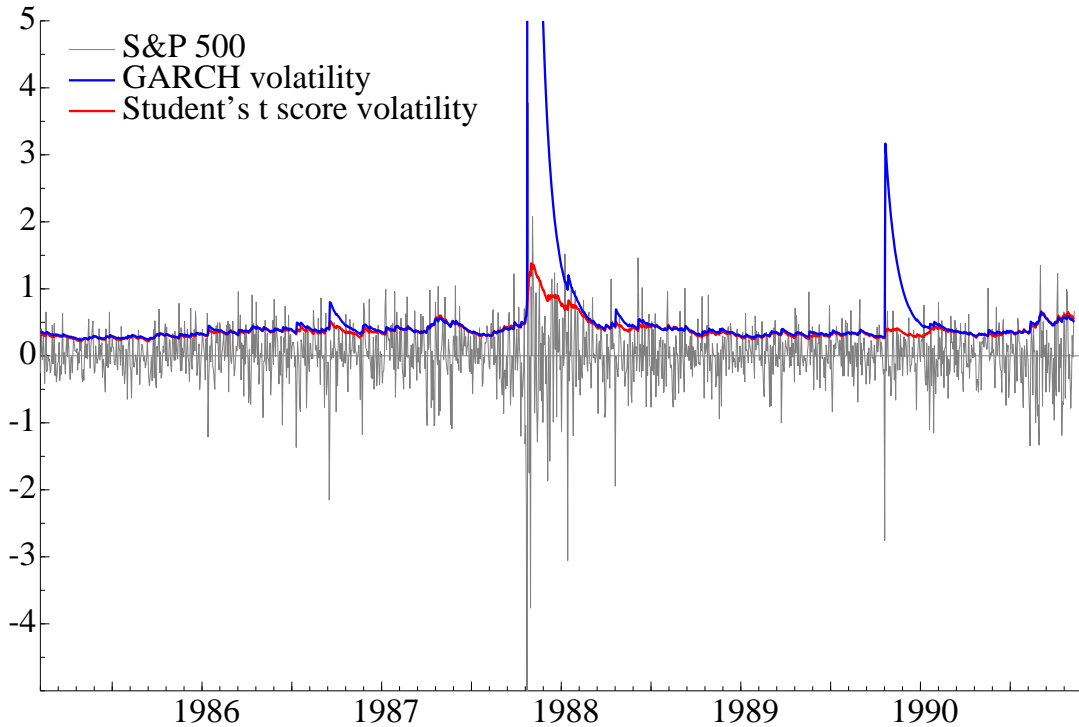


Figure 1: Typical differences between GARCH and Student’s  $t$  based score-driven volatility dynamics

by a slow exponential decline. By contrast, the filtered volatility estimate obtained from the Student’s  $t$  based score-driven dynamics recognizes that the large negative return in 1989 is incidental and is probably due to the fat-tailedness of the data. Similarly for the turmoil at the end of 1987, both volatilities increase, but again the score-driven filtered estimate is considerably more robust and does not exhibit the subsequent non-intuitive strong exponential decline.

The intuition for the behavior of the score filter is simple. Consider a large observation  $y_t$ , relative to the current level of volatility  $f_t$ . If we assume that tail observations are unlikely because  $y_t$  is conditionally Gaussian, a large value of  $y_t$  must be attributed to a (steep) volatility increase. If, however, we assume that data are conditionally fat-tailed, part of the magnitude of  $y_t$  can be attributed to the fat-tails. As a result, the volatility needs to be increased much less dramatically. This is precisely what the score-driven approach attains in this case. The GARCH dynamics do not account for this: whatever the degree of conditional fat-tailedness, the next volatility level always responds quadratically to any observation  $y_t$ . The score-driven dynamics, by contrast, incorporates the trade-off by down-weighting the impact of a large  $y_t^2$  automatically; see Creal et al. (2011, 2013) for more discussions and examples.

## 2.2 The General Setting

We next provide a more formal account of the score-driven model in order to derive our asymptotic results. A score-driven model is characterized by three specific building blocks.



These include the distributional assumption, the choice of parameterization, and the type of measurement equation. We use the following general framework throughout the remainder of this paper. Consider a  $d_y$ -dimensional stochastic sequence  $\{y_t\}_{t \in \mathbb{N}}$  given by

$$y_t = g(f_t(\boldsymbol{\theta}, \bar{f}_1), u_t), \quad u_t \sim p_u(u_t; \lambda), \quad (2.5)$$

where  $g : \mathcal{F} \times \mathcal{U} \rightarrow \mathcal{Y}$  is a link function that is strictly increasing in its second argument,  $f_t(\boldsymbol{\theta}, \bar{f}_1)$  is the time-varying parameter function with outcomes in the convex set  $\mathcal{F}$ ,  $\{u_t\}_{t \in \mathbb{N}}$  is an exogenous i.i.d. sequence of random variables for every parameter vector  $\lambda \in \Lambda \subseteq \mathbb{R}^{d_\lambda}$ , and  $p_u$  is a density function. The time-varying parameter updating scheme is given by

$$f_{t+1}(\boldsymbol{\theta}, \bar{f}_1) = \omega + \alpha s(f_t(\boldsymbol{\theta}, \bar{f}_1), y_t; \lambda) + \beta f_t(\boldsymbol{\theta}, \bar{f}_1), \quad (2.6)$$

for  $t > 1$ , and initialized at  $f_1(\boldsymbol{\theta}, \bar{f}_1) = \bar{f}_1$  for a nonrandom  $\bar{f}_1 \in \mathcal{F} \subseteq \mathbb{R}$ . The vector  $\boldsymbol{\theta}^\top = (\omega, \alpha, \beta, \lambda^\top) \in \Theta \subseteq \mathbb{R}^{3+d_\lambda}$  is the static parameter vector, and  $s : \mathcal{F} \times \mathcal{Y} \times \Lambda \rightarrow \mathcal{F}$  is the scaled score of the conditional density of  $y_t$  given  $f_t$ . For ease of exposition, we assume that  $\lambda$  is a scalar, that is  $d_\lambda = 1$ . Whenever possible, we suppress the dependence of  $f_t(\boldsymbol{\theta}, \bar{f}_1)$  on its arguments and write  $f_t$  instead.

**Main example (continued):** For the time-varying volatility model of Section 2.1, the link function  $g$  in (2.5) simplifies to  $g(f_t, u_t) = f_t^{1/2} \cdot u_t$ , which is strictly increasing in  $u_t$  if  $f_t > 0$ . For a location model, the link function is additive rather than multiplicative. Depending on the application the link function may also be much more complex, such as for the Beta distribution with time-varying parameters; see Creal et al. (2014).

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

$$p_y(y_t | f_t; \lambda) = p_u(\bar{g}(f_t, y_t); \lambda) \cdot \bar{g}'(f_t, y_t), \quad (2.7)$$

where

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

is the inverse of  $g(f_t, u_t)$  with respect to its second argument,  $u_t$ , and where  $\bar{g}'_t := \bar{g}'(f_t, y_t) := \partial \bar{g}(f_t, y) / \partial y |_{y=y_t}$  is the Jacobian of the transformation.

**Main example (continued):** For the volatility model from Section 2.1, we have  $g(f_t, u_t) = f_t^{1/2} \cdot u_t$ , and thus  $\bar{g}_t = g^{-1}(f_t, y_t) = y_t / f_t^{1/2}$  and  $\bar{g}'_t = f_t^{-1/2}$ .

The defining aspect of the score-driven model is its use of the scaled score function as the driving mechanism in transition equation (2.6). The scaled score function is defined as

$$s(f_t, y_t; \lambda) = S(f_t; \lambda) \cdot \nabla_t(f_t, y_t; \lambda) \quad \text{where} \quad \nabla_t(f_t, y_t; \lambda) = \left[ \frac{\partial \bar{p}_t}{\partial f} + \frac{\partial \bar{g}'_t}{\partial f} \right] \Big|_{f=f_t}, \quad (2.8)$$

with  $\bar{p}_t := \bar{p}(f_t, y_t; \lambda) = \log p_u(\bar{g}(f_t, y_t); \lambda)$  and where  $S : \mathcal{F} \times \Lambda \rightarrow \mathcal{F}$  is a positive scaling function; see Creal et al. (2013).

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

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

with the average log-likelihood function  $\ell_T$  given in closed form as

$$\ell_T(\boldsymbol{\theta}, \bar{f}_1) = \frac{1}{T} \sum_{t=1}^T \left( \log p_u(\bar{g}_t; \lambda) + \log \bar{g}'_t \right) = \frac{1}{T} \sum_{t=1}^T \left( \bar{p}_t + \log \bar{g}'_t \right). \quad (2.9)$$

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 led to the widespread application of GARCH models in applied empirical work. As is clear from equation (2.9), score-driven models benefit from the same computational advantages.

Before the asymptotic properties of MLE can be properly developed, we need to establish stationarity, ergodicity, invertibility and moment existence from primitive conditions which we do in Section 3.

### 3 Stochastic Properties of Score-Driven Filters

The likelihood function (2.9) is formulated in terms of the data and in terms of the filtered time-varying parameter  $f_t$  as defined by the recursion (2.6). 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  $\{f_t\}$  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  $\{f_t\}$  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. In this section we investigate the properties of both the filtered and model-implied sequences. The results developed in this section 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  $c^t \|x_t - x'_t\| \xrightarrow{a.s.} 0$  for some  $c > 1$ ; see Straumann and Mikosch (2006) for more details. Finally, we use  $x_t \perp x'_t$  to denote independence between  $x_t$  and  $x'_t$ .

Propositions 3.1 and 3.3 below are written specifically for the score-driven recursion (2.6). The propositions can, however, be extended to more general forms which can be found in Technical Appendix E. First, we consider the score-driven model defined in terms of the error terms  $u_t$  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  $s_u(f_t^u, u_t; \lambda) := s(f_t^u, g(f_t^u, u_t); \lambda)$  and let  $\{f_t^u\}_{t \in \mathbb{N}}$  be generated by

$$f_{t+1}^u = \omega + \alpha s_u(f_t^u, u_t; \lambda) + \beta f_t^u, \quad (3.1)$$

for  $t > 1$  and initial condition  $f_1^u = \bar{f}_1$ , where we use  $f_t^u$  as a shorthand for  $f_t^u = f_t^u(\boldsymbol{\theta}, \bar{f}_1)$  if no confusion is caused.

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

$$f_{t+1}^u = \omega + \left( \beta + \alpha \left( \frac{(1 + \lambda^{-1})u_t^2}{1 + \lambda^{-1}u_t^2} - 1 \right) \right) \cdot f_t^u, \quad (3.2)$$

such that  $s_u(f_t^u, u_t; \lambda) = ((1 + \lambda^{-1})u_t^2 / (1 + \lambda^{-1}u_t^2) - 1) \cdot f_t$ . So whereas the recursion in (2.4) is highly nonlinear in  $f_t$  given  $u_t$ , the recursion in (3.2) is linear in  $f_t^u$  for given  $u_t$ .

We next formulate a result for the stationarity and existence of moments of  $f_t^u$  as given by (3.1); compare the specific example in (3.2). We assume that the scaled score in  $u_t$  satisfies  $s_u \in \mathbb{C}^{(1,0,0)}(\mathcal{F}^* \times \mathcal{U} \times \Lambda)$  for some convex  $\mathcal{F} \subseteq \mathcal{F}^* \subseteq \mathbb{R}$ , i.e.,  $s_u$  is continuously differentiable in  $f_t^u$  and continuous in  $u_t$  and  $\lambda$ . Define  $\dot{s}_{u,t}(f^*; \lambda) := \partial s_u(f, u_t; \lambda) / \partial f|_{f=f^*}$  and the supremum

$$\rho_t^k(\boldsymbol{\theta}) := \sup_{f^* \in \mathcal{F}} |\beta + \alpha \dot{s}_{u,t}(f^*; \lambda)|^k. \quad (3.3)$$

We then have the following proposition.

**Proposition 3.1.** *For every  $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^{3+d_\lambda}$  let  $\{u_t\}_{t \in \mathbb{Z}}$  be an i.i.d. sequence and assume  $\exists \bar{f}_1 \in \mathcal{F}$  such that*

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

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

Then  $\{f_t^u(\boldsymbol{\theta}, \bar{f}_1)\}_{t \in \mathbb{N}}$  converges exponentially fast almost surely (e.a.s.) to the unique stationary and ergodic (SE) sequence  $\{f_t^u(\boldsymbol{\theta})\}_{t \in \mathbb{Z}}$  for every  $\boldsymbol{\theta} \in \Theta$  as  $t \rightarrow \infty$ .

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

$$(iii) \quad \|s_u(\bar{f}_1, u_1; \lambda)\|_{n_f^*} < \infty;$$

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

$$(v) \quad f_t^u(\boldsymbol{\theta}, \bar{f}_1) \perp \rho_t^{n_f^*}(\boldsymbol{\theta}) \quad \forall (t, \bar{f}_1) \in \mathbb{N} \times \mathcal{F};$$

then  $\sup_t \mathbb{E} |f_t^u(\boldsymbol{\theta}, \bar{f}_1)|^{n_f^*} < \infty$  and  $\mathbb{E} |f_t^u(\boldsymbol{\theta})|^{n_f} < \infty \quad \forall \boldsymbol{\theta} \in \Theta$  and  $n_f \in [0, n_f^*]$ .

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  $f_t^u(\boldsymbol{\theta})$  in terms of  $\{u_t\}_{t=-\infty}^{t-1}$ . We refer to the Technical Appendix for further details.

**Remark 3.2.** Independence of  $u_t$  and  $f_t^u(\boldsymbol{\theta}, \bar{f}_1)$  is sufficient to imply condition (v), i.e., if  $u_t \perp f_t^u(\boldsymbol{\theta}, \bar{f}_1) \quad \forall (t, \boldsymbol{\theta}, \bar{f}_1)$ , then condition (v) in Proposition 3.1 holds. In addition, 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  $f_t^u$ ; 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  $\lambda + 1 < \infty$ . We also have that  $\rho_1^{n_f^*}(\boldsymbol{\theta})$  and  $f_t^u$  are independent, as the former only depends on  $u_t$  and the latter on  $u_{t-1}, u_{t-2}, \dots$ . Hence condition (v) is satisfied. Conditions (ii) and (iv) are satisfied if the factor in front of (3.2) has a log-moment or an  $n_f$  moment, respectively. For example, for  $n_f^* = 1$  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  $\boldsymbol{\theta}_0 \in \Theta$ .

Irrespective of whether we have correct or 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 stated below is key in establishing the invertibility, moment bounds and e.a.s. convergence uniformly over the parameter space  $\Theta$  of the score-driven filtered sequence  $\{f_t(\boldsymbol{\theta}, \bar{f}_1)\}$ , formulated in terms of the data  $\{y_t\}$  rather than in terms of the innovations  $\{u_t\}$  as in equation (3.1). To state our subsequent proposition concisely, we define  $\dot{s}_{y,t}(f^*; \lambda) := \partial s(f, y_t; \lambda) / \partial f|_{f=f^*}$  and the supremum

$$\bar{\rho}_t^k(\boldsymbol{\theta}) = \sup_{f^* \in \mathcal{F}} |\beta + \alpha \dot{s}_{y,t}(f^*; \lambda)|^k. \quad (3.4)$$

**Proposition 3.3.** *Let  $\Theta \subset \mathbb{R}^{3+d_\lambda}$  be compact,  $s \in \mathbb{C}^{(1,0,0)}(\mathcal{F} \times \mathcal{Y} \times \Lambda)$ , and let  $\{y_t\}_{t \in \mathbb{Z}}$  be an SE sequence. Assume  $\exists \bar{f}_1 \in \mathcal{F}$  such that*

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

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

*Then uniformly on  $\Theta$  the sequence  $\{f_t(\boldsymbol{\theta}, \bar{f}_1)\}_{t \in \mathbb{N}}$  converges e.a.s. to a unique SE sequence  $\{f_t(\boldsymbol{\theta})\}_{t \in \mathbb{Z}}$ , as  $t \rightarrow \infty$ .*

*If furthermore  $\exists n_f^* > 0$  such that*

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

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

$$(v) \sup_{\boldsymbol{\theta} \in \Theta} f_t(\boldsymbol{\theta}, \bar{f}_1) \perp \sup_{\boldsymbol{\theta} \in \Theta} \bar{\rho}_t^{n_f^*}(\boldsymbol{\theta}) \quad \forall (t, \bar{f}_1);$$

*then  $\sup_t \|f_t(\cdot, \bar{f}_1)\|_{n_f^*}^\Theta < \infty$  and  $\|f_t(\cdot)\|_{n_f^*}^\Theta < \infty$  for  $n_f \in [0, n_f^*]$ .*

The conditions of Proposition 3.3 are easily satisfied by many specific models. Let us illustrate this point by turning to our main example.

**Main example (continued):** *Consider the volatility model in equation (2.4) with  $0 < \underline{\lambda} \leq \lambda \leq \bar{\lambda} < \infty$ . From the uniform boundedness of the score in  $y_t$  for given  $\bar{f}_1$ , we obtain that conditions (i) and (iii) are trivially satisfied. Furthermore, since*

$$\dot{s}_{y,t}(f^*; \lambda) = \frac{(1 + \lambda^{-1})y_t^4 / (\lambda f^{*2})}{(1 + \lambda^{-1}y_t^2 / f^*)^2}, \quad (3.5)$$

*we have that,*<sup>1</sup>

$$\sup_{f^*} \left| \beta - \alpha + \alpha \frac{(1 + \lambda^{-1})y_t^4 / (\lambda f^{*2})}{(1 + y_t^2 / (\lambda f^*))^2} \right| \leq \max(|\beta - \alpha|, |\beta + \alpha \lambda|). \quad (3.6)$$

<sup>1</sup>The supremum over  $f^*$  for fixed  $\lambda$  and  $y_t$  is reached at either  $f^* \rightarrow \infty$  (yielding the value  $\beta - \alpha$ ) or  $f^* \rightarrow 0$  (yielding the value  $\beta + \alpha \lambda$ ). Note, however, that the latter is in general unattainable as  $f_t > \min(\bar{f}_1, \omega / (1 + \alpha - \beta)) > 0$ .

Hence, using Remark 3.2 and taking the upper bound of the supremum, we can set  $\bar{\rho}_t^k(\boldsymbol{\theta}) = \max(\beta - \alpha, \beta + \lambda\alpha)^k$ , given the parameter restriction  $\beta > \alpha > 0$ . As  $\bar{\rho}_t^k(\boldsymbol{\theta})$  is independent of the data  $y_t$ , condition (v) is trivially satisfied. Moreover, conditions (ii) and (iv) simplify to

$$\sup_{\boldsymbol{\theta} \in \Theta} \max(\beta - \alpha, \beta + \lambda\alpha) < 1. \quad (3.7)$$

In practice, it may be difficult to verify the contraction condition (iv) of Proposition 3.3 when  $\bar{\rho}_t(\boldsymbol{\theta})$  depends on the unknown measure of the data. Furthermore, in such cases, the independence condition (v) is difficult to satisfy since the data sequence is not independent over time. This independence however only needs to hold for some random variable that bounds  $\bar{\rho}_1^{n_f}(\boldsymbol{\theta})$  and satisfies the contraction condition. More generally, Remark 3.4 states an alternative uniform condition that is typically available for fat tailed score-driven models and renders both conditions (iv) and (v) redundant.

**Remark 3.4.** If the uniform bound  $\sup_{(f^*, y, \boldsymbol{\theta}) \in \mathcal{F} \times \mathcal{Y} \times \Theta} |\beta + \alpha \partial s(f^*, y; \lambda) / \partial f| < 1$  holds, as in our main example, then we can drop conditions (iv) and (v) in Proposition 3.3.

Conditions (iii) and (iv) in Proposition 3.3 imply its conditions (i) and (ii), respectively. We emphasize that under conditions (i) and (ii) our score filter is *invertible* since we are able to write  $f_t(\boldsymbol{\theta})$  as a measurable function of all past observations. Most importantly, the invertibility property ensures that the effect of the initialization  $\bar{f}_1$  vanishes as  $t \rightarrow \infty$ , and that the filter converges to a unique limit process independently of  $\bar{f}_1$ ; see, for example, Granger and Andersen (1978), Straumann and Mikosch (2006), Wintenberger (2013) and Blasques et al. (2016). Establishing invertibility is usually one of the main challenges for nonlinear time series models with 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; 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 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.

In the remainder of this section we extend the results of Proposition 3.3 to the derivative processes  $\partial f_t(\boldsymbol{\theta}, \bar{f}_1) / \partial \boldsymbol{\theta}$  and  $\partial^2 f_t(\boldsymbol{\theta}, \bar{f}_1) / \partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top$ . 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  $\mathbf{f}_t^{(i)}(\boldsymbol{\theta}, \bar{\mathbf{f}}_1^{0:i}) \in \mathcal{F}^{(i)}$  denote a vector containing all the  $i$ th order derivatives of  $f_t$  with respect to  $\boldsymbol{\theta}$ , where  $\bar{\mathbf{f}}_1^{0:i} \in \mathcal{F}^{(0:i)}$  contains the fixed initial condition for  $f_t$  and its derivatives up to order  $i$ . Similarly,  $\mathbf{f}_t^{(0:i)}(\boldsymbol{\theta}, \bar{\mathbf{f}}_1^{0:i}) \in \mathcal{F}^{(0:i)} = \mathcal{F} \times \dots \times \mathcal{F}^{(i)}$  denotes a vector containing  $f_t$  as well as its derivatives with respect to  $\boldsymbol{\theta}$  up to order  $i$ . Furthermore, in order to work with primitive conditions we use the notion of moment preserving maps, which we define as follows.

**Definition 3.1.** (Moment Preserving Maps)

A function  $h : \mathbb{R}^q \times \Theta \rightarrow \mathbb{R}$  is said to be  $\mathbf{n}/n$ -moment preserving, denoted as  $h(\cdot; \boldsymbol{\theta}) \in \mathbb{M}_{\Theta_1, \Theta_2}(\mathbf{n}, n)$ , if and only if  $\mathbb{E} \sup_{\boldsymbol{\theta} \in \Theta_1} |x_{i,t}(\boldsymbol{\theta})|^{n_i} < \infty$  for  $\mathbf{n} = (n_1, \dots, n_q)$  and  $i = 1, \dots, q$

implies  $\mathbb{E} \sup_{\boldsymbol{\theta} \in \Theta_2} |h(\mathbf{x}_t(\boldsymbol{\theta}); \boldsymbol{\theta})|^n < \infty$ . If  $\Theta_1$  or  $\Theta_2$  consists of a singleton, we replace  $\Theta_1$  or  $\Theta_2$  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 the 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  $f_t(\boldsymbol{\theta}, \bar{\mathbf{f}}_1)$ , and derivatives of the score, such as  $\partial s(f_t, y; \lambda)/\partial \lambda$  and  $\partial^2 s(f_t, y; \lambda)/\partial f_t \partial \lambda$ . For instance, each polynomial function  $h(x; \boldsymbol{\theta}) = \sum_{j=0}^J \theta_j x^j \forall (x, \boldsymbol{\theta}) \in \mathcal{X} \times \Theta$ ,  $\boldsymbol{\theta} = (\theta_0, \dots, \theta_J) \in \Theta \subseteq \mathbb{R}^J$  trivially satisfies  $h \in \mathbb{M}_{\boldsymbol{\theta}, \Theta}(n, m)$  with  $m = n/J \forall \boldsymbol{\theta} \in \Theta$ . If  $\Theta$  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  $h \in \mathbb{M}_{\boldsymbol{\theta}, \Theta}(n, m)$  with  $m = n/k \forall \boldsymbol{\theta} \in \Theta$ . If furthermore  $\sup_{\boldsymbol{\theta} \in \Theta} \bar{h}_k(\boldsymbol{\theta}) \leq \bar{h} < \infty$ , 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  $\mathbb{M}_{\Theta', \Theta'}(\mathbf{n}, n) \subseteq \mathbb{M}_{\Theta, \Theta}(\mathbf{n}, n^*)$  for all  $n^* \leq n$ , and all  $\Theta \subseteq \Theta'$ .

Using this notation, we let  $s \in \mathbb{M}_{\Theta, \Theta}(\mathbf{n}, n_s)$  where  $\mathbf{n} = (n_f, n_y)$ , 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  $n_s^\lambda$  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 its arguments have  $n_f$  and  $n_y$  moments. 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 partial derivatives. Finally, we use  $\bar{n}$  to denote moments of functions after the taking 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

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

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

**Proposition 3.5.** *Let the conditions of Proposition 3.3 hold with some  $n_f^* > n_f > 0$  and suppose that  $s \in \mathbb{C}^{(2,0,2)}(\mathcal{F} \times \mathcal{Y} \times \Lambda)$ . Let  $\bar{s} \in \mathbb{C}^{(2,0,2)}(\mathcal{F} \times \mathcal{Y} \times \Lambda)$  denote the scaled score evaluated at  $\bar{\mathbf{f}}_1$ , i.e.,  $s(\bar{\mathbf{f}}_1, y_t; \lambda)$ .*

*Let  $\min\{n_s, n_s^\lambda, \bar{n}_s^f, \bar{n}_s^{\lambda f}, \bar{n}_s^{ff}\} > 0$ . Then  $\{\mathbf{f}_t^{(1)}(\boldsymbol{\theta}, \bar{\mathbf{f}}_1^{0:1})\}_{t \in \mathbb{N}}$  converges e.a.s. to a unique SE sequence  $\{\mathbf{f}_t^{(1)}(\boldsymbol{\theta})\}_{t \in \mathbb{Z}}$ , uniformly in  $\Theta$ , and furthermore, we have  $\sup_t \|\mathbf{f}_t^{(1)}(\cdot, \bar{\mathbf{f}}_1^{0:1})\|_{n_{f\boldsymbol{\theta}}^*}^\Theta < \infty$  and  $\|\mathbf{f}_t^{(1)}(\cdot)\|_{n_{f\boldsymbol{\theta}}^*}^\Theta < \infty$  for any  $n_{f\boldsymbol{\theta}} \in [0, n_{f\boldsymbol{\theta}}^*]$ , where*

$$n_{f\boldsymbol{\theta}}^* := \min \{n_f, n_s, n_s^\lambda\}.$$

*If additionally  $\min\{n_s^{\lambda\lambda}, \bar{n}_s^{\lambda\lambda f}, \bar{n}_s^{\lambda f f}, \bar{n}_s^{f f f}\} > 0$ , then it follows that the second derivative process  $\{\mathbf{f}_t^{(2)}(\boldsymbol{\theta}, \bar{\mathbf{f}}_1^{0:2})\}_{t \in \mathbb{N}}$  converges e.a.s. to a unique SE sequence  $\{\mathbf{f}_t^{(2)}(\boldsymbol{\theta})\}_{t \in \mathbb{Z}}$ , uniformly in  $\Theta$ . Furthermore,  $\sup_t \|\mathbf{f}_t^{(2)}(\cdot, \bar{\mathbf{f}}_1^{0:2})\|_{n_{f\boldsymbol{\theta}}^*}^\Theta < \infty$  and  $\|\mathbf{f}_t^{(2)}(\cdot)\|_{n_{f\boldsymbol{\theta}}^*}^\Theta < \infty$  for any  $n_{f\boldsymbol{\theta}} \in [0, n_{f\boldsymbol{\theta}}^*]$ , where*

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

The expressions for  $n_{f_\theta}^*$  and  $n_{f_{\theta\theta}}^*$  may appear complex at first sight. However, they arise naturally from expressions for the derivative of  $f_t(\boldsymbol{\theta}, \bar{f}_1)$  with respect to  $\boldsymbol{\theta}$ . Let us turn to our main example and analyze the moment conditions of Proposition 3.5 in a practical setting.

**Main example (continued):** From Proposition 3.3, the limit filtered process  $f_t(\boldsymbol{\theta})$  is guaranteed to have  $n_f$  moments, uniformly in  $\boldsymbol{\theta}$ , as long as the contraction condition is satisfied. Since  $n_f < n_f^*$ , where  $n_f^*$  is determined via condition (iii) of Proposition 3.3, we can set  $n_f^*$  arbitrarily high and hence, since  $s_t$  is uniformly bounded in  $y_t$  for fixed  $\bar{f}_1$ , we can set  $n_f > 0$ . The remaining derivatives are straightforward to check and can be found in the Technical Appendix. From these expressions, we obtain  $n_s, n_s^\lambda \leq n_f$ , and  $\bar{n}_s^f, \bar{n}_s^{\lambda f}, \bar{n}_s^{ff} \rightarrow \infty$ , such that  $\min\{n_s, n_s^\lambda, \bar{n}_s^f, \bar{n}_s^{\lambda f}, \bar{n}_s^{ff}\} > 0$  and  $n_{f_\theta} < n_{f_\theta}^* = n_f$ . We also obtain  $n_s^{\lambda\lambda} \leq n_f$  and  $n_s^f, n_s^{ff}, n_s^{f\lambda}, \bar{n}_s^{\lambda\lambda f}, \bar{n}_s^{\lambda ff}, \bar{n}_s^{fff} \rightarrow \infty$ , such that  $\min\{n_s^{\lambda\lambda}, \bar{n}_s^{\lambda\lambda f}, \bar{n}_s^{\lambda ff}, \bar{n}_s^{fff}\} > 0$  and  $n_{f_{\theta\theta}} < n_{f_{\theta\theta}}^* = \min\{n_{f_\theta}, n_f, n_{f_\theta}/2\} = n_f/2$ . 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 volatility 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 F provides a detailed compendium of the moment preserving properties of different classes of functions. With this compendium, the verification of primitive moment conditions in Proposition 3.5 is considerably simplified for many relevant settings of practical interest.

## 4 Global Identification, Consistency, and A. Normality

We now formulate the 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 start with two rather standard assumptions.

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

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

The conditions in Assumption 4.2 are sufficient for the scaled score to be twice continuously differentiable, i.e.,  $s \in \mathbb{C}^{(2,0,2)}(\mathcal{F} \times \mathcal{Y} \times \Lambda)$ . Let  $\Xi$  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  $\hat{\boldsymbol{\theta}}_T(\bar{f}_1) : \Xi \rightarrow \Theta$  satisfying  $\hat{\boldsymbol{\theta}}_T(\bar{f}_1) \in \arg \max_{\boldsymbol{\theta} \in \Theta} \ell_T(\boldsymbol{\theta}, \bar{f}_1)$ , for all  $T \in \mathbb{N}$  and every initialization  $\bar{f}_1 \in \mathcal{F}$ .

Using our notation for moment-preserving maps, let  $\log \bar{g}' \in \mathbb{M}_{\Theta, \Theta}(\mathbf{n}, n_{\log \bar{g}'})$  and  $\bar{p} \in \mathbb{M}_{\Theta, \Theta}(\mathbf{n}, n_{\bar{p}})$  as defined below (2.7) and (2.8), respectively, where  $\mathbf{n} := (n_f, n_y)$ . Similarly, we have denoted  $\nabla_t$  as the unscaled score  $\partial \log p_y(y_t | f_t; \lambda) / \partial f_t$  and we let  $\sup_f |\nabla_t| \in \mathbb{M}_{\Theta, \Theta}(\mathbf{n}, \bar{n}_{\nabla})$  where  $\bar{n}_{\nabla}$  denotes the moments of  $\sup_f |\nabla_t|$ .

To establish consistency, we use the following two assumptions.

**Assumption 4.4.**  $\exists \Theta^* \subseteq \mathbb{R}^{3+d_\lambda}$ ,  $n_f > 0$  and  $\delta > 0$  such that, for every  $\bar{f}_1 \in \mathcal{F}$ ,

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

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

**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  $\{f_t(\boldsymbol{\theta}, \bar{f}_1)\}$  to an SE limit  $\{f_t(\boldsymbol{\theta})\}$  with  $n_f$  moments. Alternative primitive conditions leading to the same result can be found in Proposition 3.3 and the subsequent remarks. 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  $\nabla_t$ , and the scaled score  $s_t$ . The number of bounded moments of  $\bar{p}$ ,  $\log \bar{g}'$ ,  $\nabla_t$  and  $s_t$  can be easily determined as we have set out in Technical Appendix F. We illustrate the verification of these assumptions using our main example.

**Main example (continued):** From the derivations around equation (3.7), we have learned that the conditions of Assumption 4.4 can be easily satisfied for an appropriate compact parameter space  $\Theta^*$ . For Assumption 4.5, we notice that  $\bar{g}'(f_t, y_t) = 0.5f_t^{-1}$ , and hence  $n_{\log \bar{g}'} \rightarrow \infty$  given that  $f_t \geq \omega > 0$  under the parameter constraint  $\beta > \alpha > 0$  and the initialization  $\bar{f}_1 \geq \omega$ . Using the expression

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

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  $\hat{\boldsymbol{\theta}}_T(\bar{f}_1)$ .

**Theorem 4.6.** (Consistency under possible model mis-specification) *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  $\boldsymbol{\theta}_0 \in \Theta$  be the unique maximizer of the limit log-likelihood  $\ell_\infty(\cdot)$  on the parameter space  $\Theta \subseteq \Theta^*$  with  $\Theta^*$  as introduced in Assumption 4.4. Then the MLE satisfies  $\hat{\boldsymbol{\theta}}_T(\bar{f}_1) \xrightarrow{a.s.} \boldsymbol{\theta}_0$  as  $T \rightarrow \infty$  for every  $\bar{f}_1 \in \mathcal{F}$ .*



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  $\Theta$ , 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  $\theta_0$ .

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 mis-specified model setting. In particular, consistency of the MLE is obtained with respect to a pseudo-true parameter  $\theta_0 \in \Theta$  that is *assumed* to be the *unique* maximizer of the limit log-likelihood  $\ell_\infty(\theta)$ . This pseudo-true parameter minimizes the Kullback-Leibler divergence between the probability measure of  $\{y_t\}_{t \in \mathbb{Z}}$  and the measure implied by the model. The result naturally requires regularity conditions on the observed data  $\{y_t\}_{t=1}^T \subset \{y_t\}_{t \in \mathbb{Z}}$  that is generated by an unknown data generating process. Such conditions in this general setting can only be imposed by means of direct assumption. However, under an axiom of correct specification, we can 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  $\theta_0$  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 have 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. Below,  $\Lambda_*$  denotes the orthogonal projection of a set  $\Theta_* \subseteq \mathbb{R}^{3+d_\lambda}$  onto the subspace  $\mathbb{R}^{d_\lambda}$  holding the static parameters in  $\lambda$ . Furthermore, statements for almost every (f.a.e.) element in a set hold with respect to Lebesgue measure. Finally, we let  $g \in \mathbb{M}_{\Theta, \Theta}(\mathbf{n}, n_g)$  with  $\mathbf{n} = (n_f, n_u)$ , 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  $n_f$  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 4.7.**  $\exists \Theta_* \subseteq \mathbb{R}^{3+d_\lambda}$  and  $n_u \geq 0$  such that

- (i)  $\mathcal{U}$  contains an open set for every  $\lambda \in \Lambda_*$ ;
- (ii)  $\sup_{\lambda \in \Lambda_*} \mathbb{E}|u_t|^{n_u} < \infty$  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  $\lambda = \lambda'$ .

Conditions (i) and (iii) of Assumption 4.7 ensure that 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  $y_t(\boldsymbol{\theta}_0)$  up to order  $n_y$  follow from moments of  $u_t$  and  $f_t$  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  $(f, \lambda)$ . 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  $t$  volatility model, the domain of  $u_t$  is always  $\mathbb{R}$ , which satisfies part (i) of Assumption 4.7. As  $g(f, u) = f^{1/2}u$ , we can use a standard Hölder inequality to obtain  $n_g = n_f \cdot n_u / (2n_u + n_f)$ , such that parts (ii), (iii), and (iv) are satisfied for  $n_f > 0$ ,  $0 < n_u < \inf_{\Lambda^*} \lambda$ , and  $\omega > 0$ ,  $f_1 > 0$ , and  $\beta > \alpha > 0$ . Note that  $n_f$  follows from Proposition 3.1, part (iii), and can be set arbitrarily high.

**Assumption 4.8.**  $\exists \Theta_* \subseteq \mathbb{R}^{3+d_\lambda}$ ,  $n_f > 0$  and  $\delta > 0$ , such that for every  $\boldsymbol{\theta} \in \Theta_*$  and every  $\bar{f}_1 \in \mathcal{F}_s$

$$(i) \quad \|s_u(\bar{f}_1, u_1; \lambda)\|_{n_f + \delta} < \infty;$$

$$(ii) \quad \mathbb{E} \rho_t^{n_f + \delta}(\boldsymbol{\theta}) < 1;$$

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 4.8 ensure that the true sequence  $\{f_t(\boldsymbol{\theta}_0)\}$  is SE and has  $n_f$  moments by the application of Proposition 3.1 and Remark 3.2. Together with condition (iii) in Assumption 4.7 we then obtain that the data  $\{y_t(\boldsymbol{\theta}_0)\}_{t \in \mathbb{Z}}$  itself is SE and has  $n_y$  moments. The inequality stated in (4.1) in Assumption 4.8 and the assumption that  $\alpha \neq 0$  together ensure that the data  $\{y_t(\boldsymbol{\theta}_0)\}$  entering the update equation (2.6) render the filtered sequence  $\{f_t\}$  stochastic and non-degenerate.

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  $f_t$  and a term that is uniformly bounded in  $u_t$ . Hence, (i) in Assumption 4.8 is satisfied for arbitrary  $n_f$ . Furthermore, by the linearity of  $s_u$  in  $f_t$ , condition (ii) of Assumption 4.8 collapses to

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

In particular, for  $n_f = 1$ , we obtain the requirement  $|\beta| < 1$ , which together with the parameter restrictions to ensure positivity of  $f_t$  result in  $1 > \beta > \alpha > 0$ . Larger regions can be obtained for smaller values of  $n_f$ .

**Theorem 4.9** (Global Identification for correctly specified models). *Let Assumptions 4.1, 4.2, 4.4, 4.5, 4.7, and 4.8 hold and let the observed data be a subset of the realized path of a stochastic process  $\{y_t(\boldsymbol{\theta}_0)\}_{t \in \mathbb{Z}}$  generated by a 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 4.9. 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 4.10.** (Consistency for correctly specified models) *Let Assumptions 4.1, 4.2, 4.4, 4.5, 4.7, and 4.8 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  $\Theta^*$  and  $\Theta_*$  defined in Assumptions 4.4, 4.7 and 4.8. Then the MLE  $\hat{\boldsymbol{\theta}}_T(\bar{f}_1)$  satisfies  $\hat{\boldsymbol{\theta}}_T(\bar{f}_1) \xrightarrow{a.s.} \boldsymbol{\theta}_0$  as  $T \rightarrow \infty$  for every  $\bar{f}_1 \in \mathcal{F}$ .*

The consistency region  $\Theta^* \cap \Theta_*$  under correct specification is a subset of the consistency region  $\Theta^*$  for the mis-specified setting. This simply reflects the fact that the axiom of correct specification alone (without parameter space restrictions) is not enough to obtain the desired moment bounds. The parameter space also has to be (further) restricted 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 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, we let quantities like  $n_{\bar{p}}^\lambda$  and  $n_{\bar{p}}^{f\lambda}$  denote the number of bounded moments of the derivative of  $\bar{p}$  with respect to  $\lambda$  and the cross derivative with respect to  $f$  and  $\lambda$ , respectively. We also let  $n_{f_\theta}$  and  $n_{f_{\theta\theta}}$  be defined as in Proposition 3.5. Finally, for notational simplicity we define the following quantities,

$$n^* = \min \{n_\nabla, \bar{n}_\nabla^f, \bar{n}_\nabla^\lambda, \bar{n}_\nabla^{ff}, \bar{n}_\nabla^{\lambda f}, \bar{n}_\nabla^{\lambda\lambda}\}, \quad (4.2)$$

$$n_{\ell'} = \min \left\{ n_{\bar{p}}^\lambda, \frac{n_\nabla n_{f_\theta}}{n_\nabla + n_{f_\theta}} \right\}, \quad (4.3)$$

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

**Assumption 4.11.**  $\exists \Theta_*^* \subseteq \mathbb{R}^{3+d_\lambda}$  such that  $n^* > 0$ ,  $n_{\ell''} \geq 1$  and  $n_{\ell'} > 2 + \delta$  for some  $\delta > 0$ .

Similar to the moment conditions in Proposition 3.3, the moment conditions in Assumption 4.11 relate directly to low-level (primitive) elements of the model. The expressions in (4.2), (4.3) and (4.5) follow directly from the formulas for the derivatives of the log-likelihood with respect to  $\boldsymbol{\theta}$ . Having  $n_{\ell'} > 2 + \delta$  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  $n^* > 0$  is designed to ensure that the e.a.s. convergence of the filter  $f_t(\boldsymbol{\theta}, \bar{f}_1)$  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 4.12.** Let  $n$  denote the lowest of the primitive derivative moment numbers  $n_{\bar{p}}^{\lambda}$ ,  $n_{\nabla}$ , etc. Then  $n \geq 4 + \delta'$  implies  $n_{\ell'} > 2 + \delta$  and  $n_{\ell''} \geq 1$ , for some positive  $\delta'$  and  $\delta$ .

It is often just as easy, however, to check the moment conditions formulated in Assumption 4.11 directly rather than the simplified conditions in Remark 4.12. We illustrate this using our main example.

**Main example (continued):.** *For the Student's  $t$  volatility model, an 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  $n_{\bar{p}}^{\lambda\lambda}, \bar{n}_{\nabla}, \bar{n}_{\nabla}^f, \bar{n}_{\nabla}^{\lambda}, \bar{n}_{\nabla}^{\lambda\lambda}, \bar{n}_{\nabla}^{\lambda f}, \bar{n}_{\nabla}^{ff} \rightarrow \infty$ . 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,  $n^* > 0$ ,  $n_{\ell'} \leq \min\{n_y/\delta', n_{f_{\theta}}\}$  for arbitrary  $\delta' > 0$ , and  $n_{\ell''} \leq \min\{n_{f_{\theta\theta}}, n_{f_{\theta}}, \frac{1}{2}n_{f_{\theta}}\}$ . We have derived earlier that  $n_{f_{\theta\theta}} < n_{f_{\theta}}/2$ , such that  $n_{\ell'} > 2 + \delta$  and  $n_{\ell''} \geq 1$  both imply  $n_{f_{\theta}} > 2 + \delta$ . If the contraction condition is met over the entire parameter space, then as also shown earlier we can set  $n_{f_{\theta}}$  arbitrarily high and thus satisfy Assumption 4.11.*

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  $\ell_T(\boldsymbol{\theta}, \bar{f}_1)$  with respect to  $\boldsymbol{\theta}$  and evaluated at the MLE. As noted in White (1994), in the presence of dynamic mis-specification, 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 4.13 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 4.13 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 4.13.**  $\partial \bar{p}_t/\partial f$  and  $\partial \log \bar{g}'_t/\partial f$  are uniformly bounded random variables and  $\partial \bar{p}_t/\partial \lambda$  is a.s. Lipschitz continuous in  $(y_t, f_t)$ .

**Main example (continued):.** *Using the Student's  $t$  volatility 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  $\partial \log \bar{g}'_t/\partial f = 0.5f_t^{-1}$  is trivially uniformly bounded. Furthermore,  $\bar{p}_t$  is clearly Lipschitz continuous in  $(y_t, f_t)$ . Hence Assumption 4.13 holds for the leading example and asymptotic normality applies.*

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

**Theorem 4.14.** (Asymptotic normality under possible model mis-specification) *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, 4.11 and 4.13 are satisfied. Finally, let  $\boldsymbol{\theta}_0 \in \text{int}(\Theta)$  be the unique maximizer of  $\ell_\infty(\boldsymbol{\theta})$  on  $\Theta$ , where  $\Theta \subseteq \Theta^* \cap \Theta_*^*$  with  $\Theta^*$  and  $\Theta_*^*$  as defined in Assumptions 4.4 and 4.11. Then, for every  $\bar{f} \in \mathcal{F}$ , the MLE  $\hat{\boldsymbol{\theta}}_T(\bar{f}_1)$  satisfies*

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

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

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

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. Similarly, we can also drop Assumption 4.13, which was used to ensure that the ML score was NED. Finally, we no longer need to restrict ourselves to the uniform contraction condition in Assumption 4.4 which guaranteed the NED property for the score filter. 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 4.15.** (Asymptotic normality under correct specification) *Let Assumptions 4.1, 4.2, 4.4, 4.5, 4.7, 4.8, and 4.11 hold and assume  $\{y_t(\boldsymbol{\theta}_0)\}_{t \in \mathbb{Z}}$  is a random sequence generated by a score-driven model under some  $\boldsymbol{\theta}_0 \in \text{int}(\Theta)$  where  $\Theta \subseteq \Theta^* \cap \Theta_*^*$  with  $\Theta^*$ ,  $\Theta_*^*$  and  $\Theta_*^*$  defined in Assumptions 4.4, 4.7, 4.8, and 4.11. Then, for every  $\bar{f}_1 \in \mathcal{F}$ , the MLE  $\hat{\boldsymbol{\theta}}_T(\bar{f}_1)$  satisfies*

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

where  $\mathcal{I}(\boldsymbol{\theta}_0)$  is the Fisher information matrix as defined in Theorem 4.14.

Theorem 4.15 does not have a separate  $n_y$ -moment condition like Theorems 4.6 and 4.14. 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 4.7 and 4.8.

**Main example (continued):.** *To verify the conditions of Theorem 4.15 for the main example, we have already shown that 4.11 requires  $n_f > 2 + \delta$ . Using the derivations below Proposition 3.3, we subsequently showed that this condition is met if the contraction condition (3.7) is satisfied, and if some arbitrarily small moment  $n_y > 0$  of  $y_t$  exists. Given the specification  $g(f_t, u_t) = f_t^{1/2} u_t$ , the latter is ensured if  $\inf_\Lambda \lambda = \underline{\lambda} > 0$  such that an arbitrarily small moment exists for  $u_t$ , and (using Proposition 3.1) if*

$$\sup_{\Theta} \mathbb{E} \left| \beta - \alpha + \alpha \frac{(1 + \lambda^{-1})u_t^2}{1 + \lambda^{-1}u_t^2} \right|^{n_f} < 1, \quad (4.5)$$

for some small  $n_f$ , where  $u_t$  has a Student's  $t$  distribution with  $\lambda$  degrees of freedom. The condition can easily be checked numerically and ensures that a small moment  $n_f$  exists for  $f_t(\boldsymbol{\theta}_0)$ , for any  $\boldsymbol{\theta}_0$  in the compact parameter space  $\Theta$ . Together, this ensures  $n_y > 0$  under correct specification via a standard Hölder inequality. The moment conditions for the innovations  $u_t$  are thus considerably weaker than in the GARCH case. In particular, the innovations  $u_t$  can even have no integer moments, while the asymptotic normality of  $\boldsymbol{\theta}$  for the score-driven volatility model still applies.

The theorems and corollaries derived in this 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. The scope of the theory developed above can be appreciated even better by considering an additional range of empirically relevant examples that include conditional location models with fat tails, conditional location models with skewness, conditional log-volatility models, conditional duration models with fat-tailed densities, and models for non-linear transformations of location. In all these cases, the current theory can be readily applied.

## 5 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 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  $\bar{f}_1$  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.

## A Proofs of Main Results

*Proof of Proposition 3.1.* This is a special case of Proposition TA.5 in Appendix E. To see this, set

$$\phi(x_t(\boldsymbol{\theta}, \bar{x}), v_t, \boldsymbol{\theta}) = \omega + \alpha s_u(f_t(\boldsymbol{\theta}, \bar{f}_1), u_t; \lambda) + \beta f_t(\boldsymbol{\theta}, \bar{f}_1),$$

$v_t = u_t$ , and  $x_t(\boldsymbol{\theta}, \bar{x}) = f_t(\boldsymbol{\theta}, \bar{f}_1)$ . Note that  $s_u$  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  $\mathcal{X}$ . Conditions (i) and (iii)–(v) in Proposition TA.5 in Appendix E now directly follow from conditions (i) and (iii)–(v) in Proposition 3.1. Condition (ii) in Proposition TA.5 directly follows from condition (ii) in Proposition 3.1 by observing that from the mean value theorem we have

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

■

*Proof of Proposition 3.3.* The results for the sequence  $\{f_t\}$  are obtained by application of Proposition TA.8 in Appendix E with  $v_t = y_t$  and  $x_t(\boldsymbol{\theta}, \bar{x}) = f_t(\boldsymbol{\theta}, \bar{f}_1)$  and  $\phi(x_t, v_t, \boldsymbol{\theta}) = \omega + \alpha s(f_t, y_t; \lambda) + \beta f_t$ .

*Step 1, SE for  $f_t$ :* Condition (i) of Proposition TA.8 holds, because

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

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  $\Theta$ , and  $\log^+ |\bar{f}_1| < \infty$  for any  $\bar{f}_1 \in \mathcal{F} \subseteq \mathbb{R}$ , and  $\mathbb{E} \log^+ \sup_{\lambda \in \Lambda} |s(\bar{f}_1, y_t; \lambda)| < \infty$  by condition (i) in Proposition 3.3.

Condition (ii) in Proposition TA.8 holds, because

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

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.a) in Proposition TA.8 follows from condition (iv) in Proposition 3.3. Condition (iii.b) in Proposition TA.8 for  $n = n_f^*$  follows since by the  $C_r$ -inequality in (Loève, 1977, p.157), there exists a  $0 < c < \infty$  such that

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

where the last inequality follows from condition (iii) in Proposition 3.3, and where  $c < 1$  for  $0 < n_f^* < 1$ . Finally, condition (v.b) in Proposition TA.8 directly follows from condition (v) in Proposition 3.3.  $\blacksquare$

For the proof of Remark 3.4, see Technical Appendix B.

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

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

In the present context, the perturbed sequence  $\{\mathbf{f}_t^{(i)}(\boldsymbol{\theta}, \bar{\mathbf{f}}_1^{0:i})\}_{t \in \mathbb{N}}$  depends on the non-stationary sequence  $\{\mathbf{f}_t^{(0:i-1)}(\boldsymbol{\theta}, \bar{\mathbf{f}}_1^{0:i-1})\}_{t \in \mathbb{N}}$ , which is only stationary in the limit. The unperturbed recurrence is equal in all respects, except that it instead depends on the limit SE filter  $\{\mathbf{f}_t^{(0:i-1)}(\boldsymbol{\theta})\}_{t \in \mathbb{Z}}$ .

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

$$\mathbf{f}_t^{(i)}(\boldsymbol{\theta}, \bar{\mathbf{f}}_1^{0:i}) = \mathbf{A}_{i,t}^{(i)}(\boldsymbol{\theta}, \bar{\mathbf{f}}_1) + \mathbf{B}_t(\boldsymbol{\theta}, \bar{\mathbf{f}}_1) \mathbf{f}_{t-1}^{(i)}(\boldsymbol{\theta}, \bar{\mathbf{f}}_1^{0:i}), \quad (\text{A.1})$$

with  $\mathbf{B}_t(\boldsymbol{\theta}, \bar{\mathbf{f}}_1) = \beta + \alpha \partial s(f_t(\boldsymbol{\theta}, \bar{\mathbf{f}}_1), y_t; \lambda) / \partial f$  not depending on the order of the derivative  $i$ . The expressions for  $\mathbf{A}_{i,t}^{(1)}(\boldsymbol{\theta}, \bar{\mathbf{f}}_1)$  are presented in Appendix D.2 and only depend on derivatives up to order  $\mathbf{f}_t^{(i-1)}(\boldsymbol{\theta}, \bar{\mathbf{f}}_1^{0:i-1})$ . Note that both  $\mathbf{A}_{i,t}^{(1)}$  and  $\mathbf{B}_t$  are written explicitly as a function of  $\bar{\mathbf{f}}_1$  since they depend on the non-stationary filtered sequence  $f_t(\boldsymbol{\theta}, \bar{\mathbf{f}}_1)$  initialized at  $\bar{\mathbf{f}}_1$ . In contrast, we let  $\mathbf{A}_{i,t}^{(1)}(\boldsymbol{\theta})$  and  $\mathbf{B}_t(\boldsymbol{\theta})$  denote the stationary counterparts of  $\mathbf{A}_{i,t}^{(1)}(\boldsymbol{\theta}, \bar{\mathbf{f}}_1)$  and  $\mathbf{B}_t(\boldsymbol{\theta}, \bar{\mathbf{f}}_1)$ , respectively, that depend on the limit stationary filter  $f_t(\boldsymbol{\theta})$ . The recurrence convergence  $|\phi_t(\bar{x}) - \tilde{\phi}_t(\bar{x})| \xrightarrow{e.a.s.} 0$  in Straumann and Mikosch (2006) corresponds here to

<sup>2</sup>We state the convergence of  $\phi_t$  at some point  $\bar{x}$  rather than at the origin  $\phi_t(0)$  as in Straumann and Mikosch (2006) since, depending on the application, our recursion may not be well defined at  $\bar{x} = 0$ .



having  $|\mathbf{A}_{i,t}^{(1)}(\boldsymbol{\theta}, \bar{f}_1) - \mathbf{A}_{i,t}^{(1)}(\boldsymbol{\theta})| \xrightarrow{e.a.s.} 0$  and  $|\mathbf{B}_t(\boldsymbol{\theta}, \bar{f}_1) - \mathbf{B}_t(\boldsymbol{\theta})| \xrightarrow{e.a.s.} 0$ . Both conditions are easily verified. Indeed, the expressions in Appendix D.2 show that  $\mathbf{A}_{i,t}^{(1)}(\boldsymbol{\theta}, \bar{f}_1)$  satisfies

$$|\mathbf{A}_{i,t}^{(1)}(\boldsymbol{\theta}, \bar{f}_1) - \mathbf{A}_{i,t}^{(1)}(\boldsymbol{\theta})| \leq \sup_f |\partial \mathbf{A}_{i,t}^{(1)}(\boldsymbol{\theta}) / \partial f| \cdot |f_t(\boldsymbol{\theta}, \bar{f}_1) - f_t(\boldsymbol{\theta})|$$

and hence we obtain  $|\mathbf{A}_{i,t}^{(1)}(\boldsymbol{\theta}, \bar{f}_1) - \mathbf{A}_{i,t}^{(1)}(\boldsymbol{\theta})| \xrightarrow{e.a.s.} 0$  by Lemma 2.1 in Straumann and Mikosch (2006) since  $|f_t(\boldsymbol{\theta}, \bar{f}_1) - f_t(\boldsymbol{\theta})| \xrightarrow{e.a.s.} 0$  by Proposition 3.3 and  $\sup_f |\partial \mathbf{A}_{i,t}^{(1)}(\boldsymbol{\theta}) / \partial f|$  is SE with a logarithmic moment since  $\min\{n_s, n_s^\lambda, \bar{n}_s^f, \bar{n}_s^{\lambda f}\} > 0$ . Similarly, we obtain

$$|\mathbf{B}_t(\boldsymbol{\theta}, \bar{f}_1) - \mathbf{B}_t(\boldsymbol{\theta})| \leq \sup_f |\partial \mathbf{B}_t(\boldsymbol{\theta}) / \partial f| \cdot |f_t(\boldsymbol{\theta}, \bar{f}_1) - f_t(\boldsymbol{\theta})| \xrightarrow{e.a.s.} 0 \quad \text{as } t \rightarrow \infty,$$

since  $\bar{n}_s^{\lambda f} > 0$  implies that  $\sup_f |\partial \mathbf{B}_t(\boldsymbol{\theta}) / \partial f|$  is SE with a logarithmic moment, and  $|f_t(\boldsymbol{\theta}, \bar{f}_1) - f_t(\boldsymbol{\theta})|$  vanishes e.a.s. The convergence of the Lipschitz coefficients  $\sup_x |\phi'_t(x) - \phi'_t(x)| = |\mathbf{B}_t(\boldsymbol{\theta}, \bar{f}_1) - \mathbf{B}_t(\boldsymbol{\theta})| \xrightarrow{e.a.s.} 0$  follows trivially by the same argument.

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.19 of the Technical Appendix.

Finally, we note that the unperturbed recursions satisfy the conditions of Proposition TA.8. In the notation of Straumann and Mikosch (2006), this means not only that the limit recursion  $\tilde{\phi}_t$  is SE, but also, that its solution  $\{\tilde{x}_t\}_{t \in \mathbb{Z}}$  has a logarithmic moment. The logarithmic moment is obtained below.

*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.b)–(v.b) of Proposition TA.8 hold. For the limit derivative processes, we can apply Proposition TA.8 directly to the unperturbed system. For the derivative processes initialized at  $t = 1$ , we notice that the moment bounds of Proposition TA.8 can be obtained with non-stationary innovations (see Remark TA.10) as long conditions (iii.b)–(v.b) hold uniformly in  $t$ . Below, we focus on the process generated by the unperturbed system.

Inspection of the formula for  $\mathbf{A}_{i,t}^{(1)}(\boldsymbol{\theta})$  reveals that  $\mathbf{A}_{i,t}^{(1)}(\boldsymbol{\theta})$  has  $\min\{n_f, n_s, n_s^\lambda\}$  bounded moments and  $\mathbf{A}_{i,t}^{(2)}(\boldsymbol{\theta})$  has  $n_{f\theta\theta}^*$  moments as defined in Proposition 3.5. Inspection of the expression for  $\mathbf{B}_t(\boldsymbol{\theta})$  reveals that  $\mathbf{B}_t(\boldsymbol{\theta})$  has  $n_s^f$  moments.

Proposition 3.5 implies that condition (iii.b) in Proposition TA.8 holds with  $n_{f\theta}^*$  moments for the first derivative process and  $n_{f\theta\theta}^*$  moments for the second derivative process, since, for any  $n > 0$ , from the  $C_r$ -inequality in (Loève, 1977, p.157), there exists a  $0 < c < \infty$  such that,

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

Condition (iv.b) in Proposition TA.8 holds with both  $n = n_{f_\theta}^*$  and  $n = n_{f_{\theta\theta}}^*$ , since

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

because  $n_f^* > n_f \geq n_{f_\theta}^* \geq n_{f_{\theta\theta}}^*$ . Finally, condition (v.b) directly implies condition (v.b) in Proposition TA.8. We thus obtain, by Proposition TA.8,  $n_{f_\theta}^*$  ( $n_{f_{\theta\theta}}^*$ ) moments for the first (second) derivative process, initialized at  $t = 1$ , and generated by the unperturbed system, as well as,  $n_{f_\theta}$  ( $n_{f_{\theta\theta}}$ ) moments for the limit process, for any  $n_{f_\theta} < n_{f_\theta}^*$  ( $n_{f_{\theta\theta}} < n_{f_{\theta\theta}}^*$ ). ■

*Proof of Theorem 4.3.* The result follows immediately from the differentiability of  $\bar{p}$ ,  $\bar{g}$ ,  $\bar{g}'$ , the compactness of  $\Theta$ , 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  $\hat{\boldsymbol{\theta}}_T(\bar{f}_1) \xrightarrow{a.s.} \boldsymbol{\theta}_0$  from the uniform convergence of the criterion function and the identifiable uniqueness of the maximizer  $\boldsymbol{\theta}_0 \in \Theta$ ,

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

*Step 1, uniform convergence:* Let  $\ell_T(\boldsymbol{\theta})$  denote the likelihood  $\ell_T(\boldsymbol{\theta}, \bar{f}_1)$  with  $f_t(\boldsymbol{\theta}, \bar{f}_1)$  replaced by  $f_t(\boldsymbol{\theta})$ . Also define  $\ell_\infty(\boldsymbol{\theta}) = \mathbb{E} \tilde{\ell}_t(\boldsymbol{\theta}) \quad \forall \boldsymbol{\theta} \in \Theta$ , with  $\tilde{\ell}_t$  denoting the contribution of the  $t$ -th observation to the likelihood function  $\ell_T$ . We have

$$\begin{aligned}
&\sup_{\boldsymbol{\theta} \in \Theta} |\ell_T(\boldsymbol{\theta}, \bar{f}_1) - \ell_\infty(\boldsymbol{\theta})| \leq \\
&\sup_{\boldsymbol{\theta} \in \Theta} |\ell_T(\boldsymbol{\theta}, \bar{f}_1) - \ell_T(\boldsymbol{\theta})| + \sup_{\boldsymbol{\theta} \in \Theta} |\ell_T(\boldsymbol{\theta}) - \ell_\infty(\boldsymbol{\theta})|. \tag{A.2}
\end{aligned}$$

The first term vanishes by application of Lemma 2.1 in Straumann and Mikosch (2006) since  $f_t(\boldsymbol{\theta}, \bar{f}_1)$  converges e.a.s. to  $f_t(\boldsymbol{\theta})$  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.1 and TA.2 from 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  $\Theta$ , 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  $\ell_T$  in  $\boldsymbol{\theta} \in \Theta \quad \forall T \in \mathbb{N}$  and the uniform convergence of the objective function proved earlier. ■

*Proof of Theorem 4.9.* We index the true  $\{f_t\}$  and the observed random sequence  $\{y_t\}$  by the parameter  $\theta_0$ , e.g.  $\{y_t(\theta_0)\}$ , 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  $\theta_0 \in \Theta$ . As conditions (i) and (ii) of Proposition 3.1 hold immediately by Assumption 4.8 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  $\{f_t(\theta_0)\}$  is SE and has at least  $n_f$  moments for any  $\theta \in \Theta$ . The SE nature and  $n_f$  moments of  $\{f_t(\theta_0)\}$  together with part (iii) of Assumption 4.7 imply, in turn, that  $\{y_t(\theta_0)\}$  is SE with  $n_y = n_g$  moments.

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

$$Q_\infty(\theta) = \mathbb{E} \tilde{\ell}_t(\theta) = \mathbb{E} \log p_{y_t|y_{t-1}, y_{t-2}, \dots} \left( y_t(\theta) \middle| y_{t-1}(\theta_0), y_{t-2}(\theta_0), \dots; \theta \right).$$

As a normalization, we subtract the constant  $Q_\infty(\theta_0)$  from  $Q_\infty(\theta)$  and focus on showing that

$$Q_\infty(\theta) - Q_\infty(\theta_0) < 0 \quad \forall (\theta_0, \theta) \in \Theta \times \Theta : \theta \neq \theta_0.$$

To do this, we use Lemma TA.3 from Technical Appendix B and rewrite

$$Q_\infty(\theta) - Q_\infty(\theta_0) = \int \int \left[ \int p_y(y|f, \lambda_0) \log \frac{p_y(y|\tilde{f}; \lambda)}{p_y(y|f; \lambda_0)} dy \right] p_{f_t, \tilde{f}_t}(f, \tilde{f}; \theta_0, \theta) df d\tilde{f}, \quad (\text{A.3})$$

for all  $(\theta_0, \theta) \in \Theta \times \Theta : \theta \neq \theta_0$ , where  $p_{f_t, \tilde{f}_t}(f, \tilde{f}; \theta_0, \theta)$  is the bivariate pdf for the pair  $(f_t(\theta_0), \tilde{f}_t(\theta))$ . We note that the pdf  $p_{f_t, \tilde{f}_t}(f, \tilde{f}; \theta_0, \theta)$  depends on both  $\theta_0$  and  $\theta$ , as for instance the recursion defining  $\tilde{f}_t(\theta)$  depends on both  $\theta$  and on  $y_t(\theta_0)$ , which in turn depends on  $\theta_0$ . Next, we use Gibb's inequality to show that this quantity is negative for  $\theta \neq \theta_0$ .

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

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

with strict equality holding if and only if  $p_y(y|\tilde{f}; \lambda) = p_y(y|f; \lambda_0)$  almost everywhere in  $\mathcal{Y}$  w.r.t.  $p_y(y|f, \lambda_0)$ . By Lemma TA.4 from Technical Appendix B there exists a set  $Y\tilde{F}\tilde{F} \subseteq \mathcal{Y} \times \mathcal{F} \times \tilde{\mathcal{F}}$  with positive probability mass and with orthogonal projections  $Y\tilde{F} \subseteq \mathcal{Y} \times \mathcal{F}$ ,  $\tilde{F}\tilde{F} \subseteq \mathcal{F} \times \tilde{\mathcal{F}}$ , etc., for which (i)–(ii) hold if  $\lambda \neq \lambda_0$ , and for which (i)–(iii) hold if  $\lambda = \lambda_0$ , where

- (i)  $p_y(y|f, \lambda_0) > 0 \quad \forall (y, f) \in Y\tilde{F}$ ;
- (ii) if  $(\tilde{f}, \lambda) \neq (f, \lambda_0)$ , then  $p_y(y|\tilde{f}; \lambda) \neq p_y(y|f; \lambda_0) \quad \forall (y, f, \tilde{f}) \in Y\tilde{F}\tilde{F}$ ;
- (iii) if  $\lambda = \lambda_0$  and  $(\omega, \alpha, \beta) \neq (\omega_0, \alpha_0, \beta_0)$ , then  $f \neq \tilde{f}$  for every  $(f, \tilde{f}) \in \tilde{F}\tilde{F}$ .

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  $Y\tilde{F}\tilde{F}$  has positive probability mass. If  $\lambda = \lambda_0$ , property (iii) ensures  $f \neq \tilde{f}$  on a subset  $\tilde{F}\tilde{F}$  with positive probability mass, and hence the strict inequality again follows via (ii) and (i).  $\blacksquare$

*Proof of Corollary 4.10.* The desired result is obtained by showing (i) that under the maintained assumptions,  $\{y_t\}_{t \in \mathbb{Z}} \equiv \{y_t(\boldsymbol{\theta}_0)\}_{t \in \mathbb{Z}}$  is an SE sequence satisfying  $\mathbb{E}|y_t(\boldsymbol{\theta}_0)|^{n_y} < \infty$ ; (ii) that  $\boldsymbol{\theta}_0 \in \Theta$  is the unique maximizer of  $\ell_\infty(\boldsymbol{\theta}, \bar{\mathbf{f}}_1)$  on  $\Theta$ ; 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 4.7 and 4.8 to ensure that  $\{f_t(\boldsymbol{\theta}_0, \bar{\mathbf{f}}_1)\}_{t \in \mathbb{N}}$  converges e.a.s. to an SE limit  $\{f_t(\boldsymbol{\theta}_0)\}_{t \in \mathbb{Z}}$  satisfying  $\mathbb{E}|f_t(\boldsymbol{\theta}_0)|^{n_f} < \infty$ . 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 4.7 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 4.9 under Assumptions 4.7 and 4.8.  $\blacksquare$

*Proof of Theorem 4.14.* Following the classical proof of asymptotic normality found e.g. in White (1994, Theorem 6.2), we obtain the desired result from:

- (i) the strong consistency of  $\hat{\boldsymbol{\theta}}_T \xrightarrow{a.s.} \boldsymbol{\theta}_0 \in \text{int}(\Theta)$ ;
- (ii) the a.s. twice continuous differentiability of  $\ell_T(\boldsymbol{\theta}, \bar{\mathbf{f}}_1)$  in  $\boldsymbol{\theta} \in \Theta$ ;
- (iii) the asymptotic normality of the score

$$\sqrt{T} \ell'_T(\boldsymbol{\theta}_0, \bar{\mathbf{f}}_1^{0:1}) \xrightarrow{d} N(0, \mathcal{J}(\boldsymbol{\theta}_0)), \quad \mathcal{J}(\boldsymbol{\theta}_0) = \mathbb{E}(\tilde{\ell}'_t(\boldsymbol{\theta}_0) \tilde{\ell}'_t(\boldsymbol{\theta}_0)^\top); \quad (\text{A.4})$$

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

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

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

*Step 1 (consistency and differentiability):* Consistency to an internal point of  $\Theta$  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  $\ell'_T(\boldsymbol{\theta}_0, \bar{\mathbf{f}}_1^{0:1})$  in (A.4) follows by applying a CLT to  $\ell'_T(\boldsymbol{\theta}_0)$ ,

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

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

$$\sqrt{T} \|\ell'_T(\boldsymbol{\theta}_0, \bar{\mathbf{f}}_1^{0:1}) - \ell'_T(\boldsymbol{\theta}_0)\| \xrightarrow{a.s.} 0 \quad \text{as } T \rightarrow \infty. \quad (\text{A.7})$$

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 mis-specification, 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.14 in Technical Appendix E 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.12 and the assumption that  $n_{\ell} \geq 2$  in Assumption 4.11. Finally, the a.s. convergence in (A.7) follows directly by Lemma TA.15 in Technical Appendix E.

*Step 3, uniform convergence of  $\ell''$ :* The proof of the uniform convergence in (iv) is similar to that of Theorem 4.3. We have

$$\sup_{\boldsymbol{\theta} \in \Theta} \|\ell''_T(\boldsymbol{\theta}, \bar{\mathbf{f}}_1^{0:2}) - \ell''_{\infty}(\boldsymbol{\theta})\| \leq \sup_{\boldsymbol{\theta} \in \Theta} \|\ell''_T(\boldsymbol{\theta}, \bar{\mathbf{f}}_1^{0:2}) - \ell''_T(\boldsymbol{\theta})\| + \sup_{\boldsymbol{\theta} \in \Theta} \|\ell''_T(\boldsymbol{\theta}) - \ell''_{\infty}(\boldsymbol{\theta})\|. \quad (\text{A.8})$$

The first term on the right-hand side of (A.8) vanishes a.s. by application of Lemma 2.1 in Straumann and Mikosch (2006) since  $\sup_{\boldsymbol{\theta} \in \Theta} \|(y_t, \mathbf{f}_t^{(0:2)}(\boldsymbol{\theta}, \bar{\mathbf{f}}_1^{0:2})) - (y_t, \mathbf{f}_t^{(0:2)}(\boldsymbol{\theta}))\| \xrightarrow{e.a.s.} 0$  and  $\sup_{\boldsymbol{\theta} \in \Theta} \sup_f \|\ell''_T(\boldsymbol{\theta})\|$  has a logarithmic moment; see Lemma TA.16 in Technical Appendix E.

The second term in (A.8) converges under a bound  $\mathbb{E} \sup_{\boldsymbol{\theta} \in \Theta} \|\tilde{\ell}''_t(\boldsymbol{\theta})\| < \infty$  by the SE nature of  $\{\ell''_T\}_{t \in \mathbb{Z}}$ . The latter is implied by continuity of  $\ell''$  on the SE sequence  $\{(y_t, \mathbf{f}_t^{(0:2)}(\cdot))\}_{t \in \mathbb{Z}}$  and Proposition 4.3 in Krenzel (1985), where SE of  $\{(y_t, \mathbf{f}_t^{(0:2)}(\cdot))\}_{t \in \mathbb{Z}}$  follows from 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 4.11 and Lemma TA.13 in Technical Appendix E.

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

*Proof of Theorem 4.15.* The desired result is obtained by applying Corollary 4.10 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  $\ell_{\infty}(\boldsymbol{\theta}, \bar{\mathbf{f}}_1)$  on  $\Theta$ . Then the statement follows along the same lines as the proof of Theorem 4.14. ■

## References

- Andres, P. and A. C. Harvey (2012). The Dynamic 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.
- Bauwens, L. and P. Giot (2000). The logarithmic acid model: an application to the bid-ask quote process of three nyse stocks. *Annales d'Economie et de Statistique* 60, 117–149.
- Billingsley, P. (1961). The lindeberg-levy theorem for martingales. *Proceedings of the American Mathematical Society* 12(5), 788–792.
- Blasques, F., P. Gorgi, S. J. Koopman, and O. Wintenberger (2016, October). Feasible Invertibility Conditions and Maximum Likelihood Estimation for Observation-Driven Models. Tinbergen Institute Discussion Papers 16-082/III, Tinbergen Institute.
- 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.
- 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.
- 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. and A. G. Billé (2017). Dynamic spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Applied Econometrics*, in press.
- 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 inflations. *Econometrica* 50, 987–1008.
- Engle, R. F. (2002). New frontiers for ARCH models. *Journal of Applied Econometrics* 17(5), 425–446.
- Engle, R. F. and J. R. Russell (1998). Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica* 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.
- Grammig, J. and K. O. Maurer (2000). Non-monotonic hazard functions and the autoregressive conditional duration model. *Econometrics Journal* 3, 16–38.
- Granger, C. and A. Andersen (1978). On the invertibility of time series models. *Stochastic Processes and their Applications* 8(1), 87 – 92.
- Hansen, P. R., A. Lunde, and J. M. Nason (2011). The model confidence set. *Econometrica* 79(2), 453–497.
- 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.
- 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.
- Jiang, J. (2010). *Large sample techniques for statistics*. Springer Science & Business Media.
- 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.
- Oh, D. H. and A. Patton (2017). Time-varying systemic risk: evidence from a dynamic copula model of cds spreads. *Journal of Business and Economic Statistics*, in press.

- Opschoor, A., P. Janus, A. Lucas, and D. van Dijk (2017). New HEAVY models for fat-tailed realized covariances and returns. *Journal of Business & Economic Statistics*, in press.
- Ord, J. K., A. Koehler, and R. D. Snyder (1997). Estimation and prediction for a class of dynamic nonlinear statistical models. *Journal of the American Statistical Association* 92(440), 1621–1629.
- 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.
- 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 heteroskedastic time series: A stochastic recurrence equations approach. *The Annals of Statistics* 34(5), 2449–2495.
- van der Vaart, A. W. (2000, June). *Asymptotic Statistics (Cambridge Series in Statistical and Probabilistic Mathematics)*. Cambridge University Press.
- White, H. (1994). *Estimation, Inference and Specification Analysis*. Cambridge Books. Cambridge University Press.
- Wintenberger, O. (2013). Continuous invertibility and stable QML estimation of the EGARCH(1, 1) model. *Scandinavian Journal of Statistics* 40(4), 846–867.



## ONLINE APPENDICES

### B Proofs of Remaining Results in the Main Paper

*Proof of Theorem 4.3.* In this theorem,  $f_t$  denotes the initialized  $f_t(\boldsymbol{\theta}, \bar{f}_1)$ . Assumption 4.2 implies that  $\ell_T(\boldsymbol{\theta}, \bar{f}_1)$  is a.s. continuous (a.s.c.) in  $\boldsymbol{\theta} \in \Theta$  through continuity of each  $\tilde{\ell}_t(\boldsymbol{\theta}, \bar{f}_1) = \ell(f_t, y, \boldsymbol{\theta})$ , ensured in turn by the differentiability of  $\bar{p}, \bar{g}, \bar{g}'$ , the implied a.s.c. of  $s(f_t, y; \lambda) = \partial \bar{p}_t / \partial f$  in  $(f_t; \lambda)$  and the resulting continuity of  $f_t$  in  $\boldsymbol{\theta}$  as a composition of  $t$  continuous maps. The compactness of  $\Theta$  implies by Weierstrass' theorem that the arg max set is non-empty a.s. and hence that  $\hat{\boldsymbol{\theta}}_T$  exists a.s.  $\forall T \in \mathbb{N}$ . Similarly, Assumption 4.2 implies that  $\ell_T(\boldsymbol{\theta}, \bar{f}_1) = \ell(\{y_t\}_{t=1}^T, \{f_t\}_{t=1}^T, \boldsymbol{\theta})$  is continuous in  $y_t \forall \boldsymbol{\theta} \in \Theta$  and hence measurable w.r.t. a Borel  $\sigma$ -algebra. The measurability of  $\hat{\boldsymbol{\theta}}_T$  follows from White (1994, Theorem 2.11) or Gallant and White (1988, Lemma 2.1, Theorem 2.2). ■

The following two lemmas support the proof of Theorem 4.6.

**Lemma TA.1.** *Under the conditions of Theorem 4.6,  $\sup_{\boldsymbol{\theta} \in \Theta} |\ell_T(\boldsymbol{\theta}, \bar{f}_1) - \ell_T(\boldsymbol{\theta})| \xrightarrow{a.s.} 0$ .*

*Proof.* The expression for the likelihood in (2.9) and the differentiability conditions in Assumption 4.2 ensure that  $\tilde{\ell}_t(\cdot, \bar{f}_1) = \ell(f_t(\cdot, \bar{f}_1), y_t, \cdot)$  is continuous in  $(f_t(\cdot, \bar{f}_1), y_t)$ . All the assumptions of Proposition 3.3 relevant for the process  $\{f_t\}$  hold as well. To see this, note that

- the compactness of  $\Theta$  is imposed in Assumption 4.1;
- the moment bound  $\mathbb{E}|y_t|^{n_y} < \infty$  is ensured in the statement of Theorem 4.6;
- the differentiability  $s \in \mathbb{C}^{(2,0,2)}(\mathcal{F} \times \mathcal{Y} \times \Lambda)$  is implied by  $\bar{g} \in \mathbb{C}^{(2,0)}(\mathcal{F} \times \mathcal{Y})$ ,  $\bar{p} \in \mathbb{C}^{(2,2)}(\tilde{\mathcal{U}} \times \Lambda)$ , and  $S \in \mathbb{C}^{(2,2)}(\mathcal{F} \times \Lambda)$ ;
- and finally, conditions (i)-(v) in Proposition 3.3 are ensured by Assumption 4.4. Note that under the alternative set of conditions proposed in Assumption 4.4, we can use Remark 3.4 and drop conditions (iv) (v) in Proposition 3.3.

As a result, there exists a unique SE sequence  $\{f_t(\cdot)\}_{t \in \mathbb{Z}}$  such that  $\sup_{\boldsymbol{\theta} \in \Theta} |f_t(\boldsymbol{\theta}, \bar{f}_1) - f_t(\boldsymbol{\theta})| \xrightarrow{a.s.} 0 \forall \bar{f}_1 \in \mathcal{F}$ , and  $\sup_t \mathbb{E} \sup_{\boldsymbol{\theta} \in \Theta} |f_t(\boldsymbol{\theta}, \bar{f}_1)|^{n_f} < \infty$  and  $\mathbb{E} \sup_{\boldsymbol{\theta} \in \Theta} |f_t(\boldsymbol{\theta})|^{n_f} < \infty$  with  $n_f$  and  $n_y$  such that  $n_\ell \geq 1$  from Assumption 4.5. Hence, the result follows immediately by an application of the continuous mapping theorem for  $\ell : \mathbb{C}(\Theta, \mathcal{F}) \times \mathcal{Y} \times \Theta \rightarrow \mathbb{R}$ . ■

**Lemma TA.2.** *Under the conditions of Theorem 4.6,  $\sup_{\boldsymbol{\theta} \in \Theta} |\ell_T(\boldsymbol{\theta}) - \ell_\infty(\boldsymbol{\theta})| \xrightarrow{a.s.} 0$ .*

*Proof.* We apply the ergodic theorem for separable Banach spaces of Rao (1962) (see also Straumann and Mikosch (2006, Theorem 2.7)) to the sequence  $\{\ell_T(\cdot)\}$  with elements taking values in  $\mathbb{C}(\Theta)$ , so that  $\sup_{\boldsymbol{\theta} \in \Theta} |\ell_T(\boldsymbol{\theta}) - \ell_\infty(\boldsymbol{\theta})| \xrightarrow{a.s.} 0$ , where  $\ell_\infty(\boldsymbol{\theta}) = \mathbb{E} \tilde{\ell}_t(\boldsymbol{\theta}) \forall \boldsymbol{\theta} \in \Theta$ . The ULLN  $\sup_{\boldsymbol{\theta} \in \Theta} |\ell_T(\boldsymbol{\theta}) - \mathbb{E} \tilde{\ell}_t(\boldsymbol{\theta})| \xrightarrow{a.s.} 0$  as  $T \rightarrow \infty$  follows, under a moment bound  $\mathbb{E} \sup_{\boldsymbol{\theta} \in \Theta} |\tilde{\ell}_t(\boldsymbol{\theta})| < \infty$ , by the SE nature of  $\{\ell_T\}_{t \in \mathbb{Z}}$ , which is implied by continuity of  $\ell$  on the SE sequence  $\{(f_t(\cdot), y_t)\}_{t \in \mathbb{Z}}$  and Proposition 4.3 in Krengel (1985). Moment bound  $\mathbb{E} \sup_{\boldsymbol{\theta} \in \Theta} |\tilde{\ell}_t(\boldsymbol{\theta})| < \infty$  is ensured by  $\sup_{\boldsymbol{\theta} \in \Theta} \mathbb{E} |f_t(\boldsymbol{\theta})|^{n_f} < \infty \forall \boldsymbol{\theta} \in \Theta$ ,  $\mathbb{E} |y_t|^{n_y} < \infty$ , and the fact that Assumption 4.5 implies  $n_\ell \geq 1$ . We stress that Assumption 4.5 can be checked via low-level conditions on  $n_y$  and  $n_f$  via the moment preserving maps as laid out in Technical Appendix F. ■

The following lemmas support the proof of Theorem 4.9.

**Lemma TA.3.** *Under the conditions of Theorem 4.9,*

$$Q_\infty(\boldsymbol{\theta}) - Q_\infty(\boldsymbol{\theta}_0) = \int \int \left[ \int p_y(y|f, \lambda_0) \log \frac{p_y(y|\tilde{f}; \lambda)}{p_y(y|\tilde{f}; \lambda_0)} dy \right] dP_{f_t, \tilde{f}_t}(f, \tilde{f}; \boldsymbol{\theta}_0, \boldsymbol{\theta}),$$

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

*Proof.* Using the observation-driven dynamic structure of the score-driven model, we can substitute the conditioning on  $\{y_s(\boldsymbol{\theta}_0)\}_{s \leq t-1}$  by the conditioning on  $f_t(\boldsymbol{\theta})$ , where  $f_t(\boldsymbol{\theta})$  is generated through the generalized autoregressive score recursion. Under the present conditions, the (non-initialized) limit process  $\{f_t(\boldsymbol{\theta})\}_{t \in \mathbb{Z}}$  is a measurable function of  $\{y_s(\boldsymbol{\theta}_0)\}_{s \leq t-1}$ , and hence SE by Krengel's theorem for any  $\boldsymbol{\theta} \in \Theta$ ; see also SM06. By substituting the conditioning, we obtain

$$\begin{aligned} Q_\infty(\boldsymbol{\theta}) - Q_\infty(\boldsymbol{\theta}_0) &= \mathbb{E} \log p_y(y_t(\boldsymbol{\theta}_0) | f_t(\boldsymbol{\theta}); \lambda) \\ &\quad - \mathbb{E} \log p_y(y_t(\boldsymbol{\theta}_0) | f_t(\boldsymbol{\theta}_0); \lambda_0) \\ &= \int \int \int \log \frac{p_y(y|\tilde{f}; \lambda)}{p_y(y|\tilde{f}; \lambda_0)} dP_{y_t, f_t, \tilde{f}_t}(y, f, \tilde{f}; \boldsymbol{\theta}_0, \boldsymbol{\theta}), \end{aligned} \tag{B.1}$$

$\forall (\boldsymbol{\theta}_0, \boldsymbol{\theta}) \in \Theta \times \Theta : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ , with  $P_{y_t, f_t, \tilde{f}_t}(y, f, \tilde{f}; \boldsymbol{\theta}_0, \boldsymbol{\theta})$  denoting the cdf of  $(y_t(\boldsymbol{\theta}_0), f_t(\boldsymbol{\theta}_0), \tilde{f}_t(\boldsymbol{\theta}))$ . Define the bivariate cdf  $P_{f_t, \tilde{f}_t}(f, \tilde{f}; \boldsymbol{\theta}_0, \boldsymbol{\theta})$  for the pair  $(f_t(\boldsymbol{\theta}_0), \tilde{f}_t(\boldsymbol{\theta}))$ . Note that this bivariate cdf depends on  $\boldsymbol{\theta}$  through the recursion defining  $\tilde{f}_t(\boldsymbol{\theta})$ , and on  $\boldsymbol{\theta}_0$  through  $y_{t-1}(\boldsymbol{\theta}_0)$  and  $f_t(\boldsymbol{\theta}_0)$ . Also note that for any  $(\boldsymbol{\theta}_0, \boldsymbol{\theta}) \in \Theta \times \Theta$  this cdf does not depend on the initialization  $\tilde{f}_1$  because, under the present conditions, the limit criterion is a function of the unique limit SE process  $\{f_t(\boldsymbol{\theta})\}_{t \in \mathbb{Z}}$ , and not of the initialized process  $\{f_t(\boldsymbol{\theta}, \tilde{f}_1)\}_{t \in \mathbb{N}}$ ; see the proof of Theorem 4.6.

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

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

where the second equality holds because under the axiom of correct specification, and conditional on  $f_t(\boldsymbol{\theta}_0)$ , observed data  $y_t(\boldsymbol{\theta}_0)$  does not depend on  $\tilde{f}_t(\boldsymbol{\theta}) \forall (\boldsymbol{\theta}_0, \boldsymbol{\theta}) \in \Theta \times \Theta : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ . We also note that the conditional distribution  $P_{y_t | f_t}(y | f, \lambda_0)$  has a density  $p_y(y | f, \lambda_0)$  defined in equation (2.7). The existence of this density follows because  $g(f, \cdot)$  is a diffeomorphism  $g(f, \cdot) \in \mathbb{D}(\mathcal{U})$  for every  $f \in \mathcal{F}$ , i.e., it is continuously differentiable and uniformly invertible with differentiable inverse.

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

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

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

**Lemma TA.4.** *Under the conditions of Theorem 4.9, for every  $\theta \neq \theta_0$  there exists a set  $Y\tilde{F}\tilde{F} \subseteq \mathcal{Y} \times \mathcal{F} \times \tilde{\mathcal{F}}$  with positive probability mass and with orthogonal projections  $Y\tilde{F} \subseteq \mathcal{Y} \times \mathcal{F}$ ,  $\tilde{F}\tilde{F} \subseteq \mathcal{F} \times \tilde{\mathcal{F}}$ , etc., for which (i)–(ii) hold if  $\lambda \neq \lambda_0$ , and for which (i)–(iii) hold if  $\lambda = \lambda_0$ , where*

$$(i) \quad p_y(y|f, \lambda_0) > 0 \quad \forall (y, f) \in Y\tilde{F};$$

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

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

*Proof.*

*Part (i):* The first result follows by noting that under the correct specification axiom, the conditional density  $p_y(y|f, \lambda_0)$  is implicitly defined by  $y_t(\theta_0) = g(f, u_t)$ ,  $u_t \sim p_u(u_t; \lambda_0)$ . Note that  $g(f, \cdot)$  is a diffeomorphism  $g(f, \cdot) \in \mathbb{D}(\mathcal{U})$  for every  $f \in \mathcal{F}_g$  and hence an open map, i.e.,  $g^{-1}(f, Y) \in \mathcal{T}(\mathcal{U}_g)$  for every  $Y \in \mathcal{T}(\mathcal{Y}_g)$  where  $\mathcal{T}(\mathbb{A})$  denotes a topology on the set  $\mathbb{A}$ . Therefore, since  $p_u(u; \lambda) > 0 \quad \forall (u, \lambda) \in U \times \Lambda$  for some open set  $U \subset \mathcal{U}$ , which exists by the assumption that  $u_t$  has a density with respect to Lebesgue measure. As a result, we obtain that there exists an open set  $Y \in \mathcal{T}(\mathcal{Y}_g)$  such that  $p_y(y|f, \lambda_0) > 0 \quad \forall (y, f) \in Y \times \mathcal{F}_g$ , namely the image of any open set  $U \subseteq \mathcal{U}$  under  $g(f, \cdot)$ . Next,  $Y\tilde{F}\tilde{F}$  can be constructed by taking the union of  $Y$  over  $\tilde{F}\tilde{F}$  for any  $\tilde{F}\tilde{F}$  of positive measure for  $\lambda \neq \lambda_0$ , and for a set  $\tilde{F}\tilde{F}$  satisfying (iii) below if  $\lambda = \lambda_0$ .

*Part (ii):* The second result is implied directly by the assumption that  $p_y(y|f, \lambda) = p_y(y|f', \lambda')$  almost everywhere in  $Y$  for some open set  $Y \subset \mathcal{Y}$  if and only if  $f = f'$  and  $\lambda = \lambda'$ . The existence of an open set  $Y$  was already argued under (i) above.

*Part (iii):* The assumptions that  $\alpha \neq 0 \quad \forall \theta \in \Theta$  (including  $\alpha_0 \neq 0$ ); and that  $\partial s(f, y; \lambda)/\partial y \neq 0$  almost everywhere in  $Y_s$  for every  $(f, \lambda) \in \mathcal{F} \times \Lambda$ ; together with the fact that  $u_t$  has a density, together ensure that both  $F$  and  $\tilde{F}$  can be chosen as open subsets, i.e., to have multiple different values.

The result is now obtained by a proof by contradiction: if  $\lambda = \lambda_0 \wedge (\omega, \alpha, \beta) \neq (\omega_0, \alpha_0, \beta_0)$ , but there is no set  $Y\tilde{F}\tilde{F}$  with positive probability mass satisfying  $f \neq \tilde{f} \quad \forall (f, \tilde{f}) \in \tilde{F}\tilde{F}$ , then it must be that  $(\omega, \alpha, \beta) = (\omega_0, \alpha_0, \beta_0)$ , which is a contradiction.

The proof goes as follows. Let  $(\theta_0, \theta) \in \Theta \times \Theta$  be a pair satisfying  $\lambda = \lambda_0 \wedge (\omega, \alpha, \beta) \neq (\omega_0, \alpha_0, \beta_0)$ . If there is no  $Y\tilde{F}\tilde{F}$  of positive probability mass with  $f \neq \tilde{f}$  for all  $(f, \tilde{f}) \in \tilde{F}\tilde{F}$ , then it must be that  $f = \tilde{f}$  except for a set of zero probability. This implies that  $\tilde{f}_t(\theta) \stackrel{a.s.}{=} f_t(\theta_0)$  for arbitrary  $t$ . Putting this into the recurrence equation for both  $f_t(\theta_0)$  and  $\tilde{f}_t(\theta)$  and subtracting the two, we obtain

$$\begin{aligned} 0 &= \phi(f_t(\theta_0), y_e, \theta) - \phi(\tilde{f}_t(\theta), y_e, \theta_0) \\ &= (\omega - \omega_0) + (\beta - \beta_0)f_t(\theta_0) + (\alpha - \alpha_0)s(f_t(\theta_0), y_t(\theta_0), \lambda_0). \end{aligned} \tag{B.2}$$

Note that  $s(f_t(\theta_0), y_t(\theta_0), \lambda_0)$  is not constant in  $y_t(\theta_0) \in Y$  where  $Y$  is an open set, because  $\alpha \neq 0 \quad \forall \theta \in \Theta$  and  $\partial s(f, y, \lambda)/\partial y \neq 0$  for every  $\lambda \in \Lambda$  and almost every  $(y, f) \in \mathcal{Y}_s \times \mathcal{F}_s$ . As a result, we must have  $\alpha = \alpha_0$  for (B.2) to hold.

Given  $\alpha = \alpha_0 \wedge \lambda = \lambda_0$ , and given  $F$  can be chosen as an open set due to the fact that  $u_t$  has a density and  $\alpha_0 > 0$ , it follows that  $\beta = \beta_0$ . Given the result for  $\alpha$  and  $\beta$ , the result  $\omega = \omega_0$  follows directly from (B.2), which establishes the contradiction and the result.  $\blacksquare$

## C Derivative Expressions for the Main Example

In this part, we provide some of the technical details of the main example of the paper, including the detailed expressions for of the required derivatives.

Let  $\{u_t\}_{t \in \mathbb{N}}$  be i.i.d. Student's  $t$  distributed noise with  $\lambda$  degrees of freedom. Consider the model  $y_t = f_t^{1/2} u_t$  as in Creal et al. (2011, 2013). Following Creal et al. (2011, 2013), we scale the score by (a time-invariant multiple of) the conditional Fisher information, which in this case amounts to setting  $S(f_t; \lambda) = 2f_t^2$ .

The following set of derivatives is straightforward (though tedious) to compute, either by hand or by a symbolic computation package such as Maple or Mathematica.

$$\begin{aligned}
\bar{p}_t &= \log \frac{\Gamma(\frac{\lambda+1}{2})}{\Gamma(\frac{\lambda}{2}) \sqrt{\pi\lambda}} - \frac{1}{2}(\lambda+1) \log \left( 1 + \frac{y_t^2}{\lambda f_t} \right), \\
\log \bar{g}'_t &= -\frac{1}{2} \log f_t \\
\nabla_t &= \frac{(1 + \lambda^{-1})y_t^2 / (2f_t^2)}{1 + y_t^2 / (\lambda f_t)} - \frac{1}{2} f_t^{-1}, \\
s_t &= \frac{(1 + \lambda^{-1})y_t^2}{1 + y_t^2 / (\lambda f_t)} - f_t, \\
\partial s_t / \partial f_t &= \frac{(1 + \lambda^{-1})y_t^4 / (\lambda f_t^2)}{(1 + y_t^2 / (\lambda f_t))^2} - 1, \\
s_{u,t} &= \left( \frac{(1 + \lambda^{-1})u_t^2}{1 + \lambda^{-1}u_t^2} - 1 \right) f_t, \\
\partial s_{u,t} / \partial f_t &= \frac{(1 + \lambda^{-1})u_t^2}{1 + \lambda^{-1}u_t^2} - 1. \\
\partial s_t / \partial \lambda &= \frac{y_t^2}{(\lambda + y_t^2 / f_t)} - \frac{(1 + \lambda)y_t^2}{(\lambda + y_t^2 / f_t)^2} = \frac{((y_t^2 / f_t)^2 - (y_t^2 / f_t))}{(\lambda + y_t^2 / f_t)^2} \cdot f_t, \\
\partial^2 s_t / \partial \lambda^2 &= \frac{-2((y_t^2 / f_t)^2 - (y_t^2 / f_t))}{(\lambda + y_t^2 / f_t)^3} \cdot f_t, \\
\partial^2 s_t / \partial \lambda \partial f_t &= \frac{(1 + \lambda)(y_t^4 / f_t^2)}{(\lambda + y_t^2 / f_t)^2} = \frac{(y_t^4 / f_t^2)}{(\lambda + y_t^2 / f_t)^2} - \frac{2(1 + \lambda)(y_t^4 / f_t^2)}{(\lambda + y_t^2 / f_t)^3}, \\
\partial^2 s_t / \partial f_t^2 &= \frac{-2(1 + \lambda^{-1})(y_t^4 / f_t^2)}{(1 + y_t^2 / (\lambda f_t))^3} \cdot \frac{1}{\lambda f_t}, \\
\partial^3 s_t / \partial \lambda^3 &= \frac{6((y_t^2 / f_t)^2 - (y_t^2 / f_t))}{(\lambda + y_t^2 / f_t)^4} \cdot f_t, \\
\partial^3 s_t / \partial \lambda^2 \partial f_t &= \frac{2(y_t^2 / f_t)^2 (\lambda + 3 - 2(y_t^2 / f_t))}{(\lambda + y_t^2 / f_t)^4}, \\
\partial^3 s_t / \partial \lambda \partial f_t^2 &= \frac{2(y_t^2 / f_t)^2 (\lambda^2 + 2\lambda - (1 + 2\lambda)(y_t^2 / f_t))}{(\lambda + y_t^2 / f_t)^4} \cdot \frac{1}{f_t}, \\
\partial^3 s_t / \partial f_t^3 &= \frac{6(1 + \lambda^{-1})y_t^4 / (\lambda f_t^2)}{(1 + y_t^2 / (\lambda f_t))^4} \cdot \frac{1}{f_t^2}.
\end{aligned}$$

We obtain directly that

- $|s_t| \leq \sup_{y_t} |s_t| < c_1 \cdot |f_t|$  for some constant  $c_1$ , and thus  $n_s \leq n_f$ .
- $\sup_{y_t} |\partial s_t / \partial \lambda| \leq c_1 \cdot |f_t|$  and thus  $n_s^\lambda \leq n_f$ .
- $\sup_{y_t} |\partial s_t / \partial f_t| \leq c_1$  and thus  $\bar{n}_s^f \rightarrow \infty$ .

- $\sup_{y_t} |\partial^2 s_t / \partial f_t \partial \lambda| \leq c_1$  and thus  $\bar{n}_s^{f\lambda} \rightarrow \infty$ .
- $\sup_{y_t} |\partial^2 s_t / \partial f_t^2| \leq c_1 f_t^{-1} \leq c_1 / \omega$  and thus  $\bar{n}_s^{ff} \rightarrow \infty$ .
- $\sup_{y_t} |\partial^2 s_t / \partial \lambda^2| \leq c_1 f_t$  and thus  $n_s^{\lambda\lambda} \leq n_f$ .
- $\sup_{y_t} |\partial^3 s_t / \partial f_t^3| \leq c_1 f_t^{-2} \leq c_1 / \omega^2$  and thus  $\bar{n}_s^{fff} \rightarrow \infty$ .
- $\sup_{y_t} |\partial^3 s_t / \partial \lambda^2 \partial f_t| \leq c_1$  and thus  $\bar{n}_s^{\lambda\lambda f} \rightarrow \infty$ .
- $\sup_{y_t} |\partial^3 s_t / \partial \lambda \partial f_t^2| \leq c_1 f_t^{-1} \leq c_1 / \omega$  and thus  $\bar{n}_s^{\lambda ff} \rightarrow \infty$ .
- $|\log \bar{g}'_t| \leq c_1 + c^2 |f_t|^\delta$  for arbitrarily small positive  $\delta$  given  $f_t > \omega$ , and thus  $n_{\log \bar{g}'} \leq n_f / \delta$ .
- $|\bar{p}_t| < c_1 + c^2 \log |1 + y_t^2 / (\lambda \omega)| \leq c_3 + c_4 |y_t|^\delta$  for arbitrarily small positive  $\delta$ , and thus  $n_{\bar{p}} \leq n_y / \delta$ .
- $|\nabla_t| < \sup_{y_t} |\nabla_t| \leq c_1 f_t^{-1} + \frac{1}{2} f_t^{-1} \leq c_2 / \omega$ , and thus  $\bar{n}_\nabla \rightarrow \infty$ .

For asymptotic normality, a further set of derivatives and moments needs to be established. We

have

$$\begin{aligned}
\bar{p}_t^\lambda &= \left( \frac{\partial}{\partial \lambda} \log \frac{\Gamma(\frac{1}{2}(\lambda+1))}{\Gamma(\lambda/2)\sqrt{\pi\lambda}} \right) + \frac{1}{2} \frac{(1+\lambda^{-1})y_t^2/(\lambda f_t)}{1+y_t^2/(\lambda f_t)} - \frac{1}{2} \log(1+y_t^2/(\lambda f_t)), \\
\bar{p}_t^{\lambda\lambda} &= -\frac{1}{2} \frac{(y_t^2/(\lambda f_t))\lambda^{-2}(2+(1-\lambda)(y_t^2/(\lambda f_t)))}{(1+y_t^2/(\lambda f_t))^2}, \\
\bar{p}_t^{\lambda f} &= -\frac{1}{2} \frac{f_t^{-1}(y_t^2/f_t)(1-y_t^2/f_t)}{(\lambda+y_t^2/f_t)^2}, \\
\bar{p}_t^{\lambda\lambda f} &= \frac{f_t^{-1}(y_t^2/f_t)(1-y_t^2/f_t)}{(\lambda+y_t^2/f_t)^3}, \\
\bar{p}_t^{\lambda f f} &= \frac{1}{2} \frac{f_t^{-2}(y_t^2/f_t)(2\lambda-3\lambda y_t^2/f_t-y_t^4/f_t^2)}{(\lambda+y_t^2/f_t)^3}, \\
\partial \nabla_t / \partial f &= \frac{\frac{1}{2}\lambda f_t^{-2}(\lambda-2\lambda y_t^2/f_t-y_t^4/f_t^2)}{(\lambda+y_t^2/f_t)^2}, \\
\partial \nabla_t / \partial \lambda &= \frac{-\frac{1}{2}f_t^{-1}(y_t^2/f_t)(1-y_t^2/f_t)}{(\lambda+y_t^2/f_t)^2}, \\
\partial^2 \nabla_t / \partial f^2 &= \frac{-\lambda f_t^{-3}(\lambda^2-3\lambda y_t^4/f_t^2-3\lambda^2 y_t^2/f_t-y_t^6/f_t^3)}{(\lambda+y_t^2/f_t)^3}, \\
\partial^2 \nabla_t / \partial f \partial \lambda &= \frac{\frac{1}{2}f_t^{-2}(y_t^2/f_t)(2\lambda-3\lambda y_t^2/f_t-y_t^4/f_t^2)}{(\lambda+y_t^2/f_t)^3}, \\
\partial^2 \nabla_t / \partial \lambda^2 &= \frac{f_t^{-1}(y_t^2/f_t)(1-y_t^2/f_t)}{(\lambda+y_t^2/f_t)^3}.
\end{aligned}$$

From this, we obtain the following.

- $|\bar{p}_t^\lambda| \leq c_1 + c_2 \log(1 + y_t/(\lambda\omega))$ , such that  $n_p^\lambda \leq n_y/\delta$  for arbitrarily small positive  $\delta$ .
- $|\bar{p}_t^{\lambda\lambda}| \leq c_1$ , such that  $n_p^{\lambda\lambda} \rightarrow \infty$ .
- $|\bar{p}_t^{\lambda f}| \leq c_1 f_t^{-1} \leq c_1/\omega$ , such that  $n_p^{\lambda f} \rightarrow \infty$ .
- $|\bar{p}_t^{\lambda\lambda f}| \leq c_1 f_t^{-1} \leq c_1/\omega$ , such that  $\bar{n}_p^{\lambda\lambda f} \rightarrow \infty$ .
- $|\bar{p}_t^{\lambda f f}| \leq c_1 f_t^{-2} \leq c_1/\omega^{-2}$ , such that  $\bar{n}_p^{\lambda f f} \rightarrow \infty$ .
- $|\partial \nabla / \partial f| \leq c_1 f_t^{-2} \leq c_1/\omega^2$ , such that  $\bar{n}_\nabla^f \rightarrow \infty$ .
- $|\partial \nabla / \partial \lambda| \leq c_1 f_t^{-1} \leq c_1/\omega$ , such that  $\bar{n}_\nabla^\lambda \rightarrow \infty$ .

- $|\partial^2 \nabla / \partial f^2| \leq c_1 f_t^{-3} \leq c_1 / \omega^3$ , such that  $\bar{n}_{\nabla}^{ff} \rightarrow \infty$ .
- $|\partial^2 \nabla / \partial f \partial \lambda| \leq c_1 f_t^{-2} \leq c_1 / \omega^2$ , such that  $\bar{n}_{\nabla}^{\lambda f} \rightarrow \infty$ .
- $|\partial^2 \nabla / \partial \lambda^2| \leq c_1 f_t^{-1} \leq c_1 / \omega$ , such that  $\bar{n}_{\nabla}^{\lambda \lambda} \rightarrow \infty$ .

## D Derivatives of the Time-Varying Parameter

### D.1 Explicit expressions for the likelihood and its derivatives

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

$$\begin{aligned}
\ell_T(\boldsymbol{\theta}, \bar{f}_1) &= \frac{1}{T} \sum_{t=1}^T \tilde{\ell}_t(\boldsymbol{\theta}, \bar{f}_1) = \frac{1}{T} \sum_{t=1}^T \ell(f_t, y_t; \lambda) \\
&= \frac{1}{T} \sum_{t=1}^T \log p_u(g^{-1}(f_t, y_t); \lambda) + \log \frac{\partial g^{-1}(f_t, y_t)}{\partial y} \\
&= \frac{1}{T} \sum_{t=1}^T \log p_u(\bar{g}_t; \lambda) + \log \frac{\partial \bar{g}_t}{\partial y} \\
&= \frac{1}{T} \sum_{t=1}^T \bar{p}_t + \log \bar{g}'_t.
\end{aligned} \tag{D.1}$$

Note that we have defined the score  $\nabla_t$  as  $\partial(\bar{p}_t + \log \bar{g}'_t) / f_t$ . The derivative of the likelihood is given by

$$\begin{aligned}
\ell'_T(\boldsymbol{\theta}, \bar{\mathbf{f}}_1^{0:1}) &= \frac{\partial \ell_T(\boldsymbol{\theta}, \bar{f}_1)}{\partial \boldsymbol{\theta}} = \sum_{t=1}^T \tilde{\ell}'_t(\boldsymbol{\theta}, \bar{\mathbf{f}}_1^{0:1}) \\
&= \frac{1}{T} \sum_{t=1}^T \frac{\partial f_t}{\partial \boldsymbol{\theta}} \cdot A_t^* + \frac{\partial \bar{p}_t}{\partial \boldsymbol{\theta}} = \frac{1}{T} \sum_{t=1}^T \frac{\partial f_t}{\partial \boldsymbol{\theta}} \cdot \nabla_t + \frac{\partial \bar{p}_t}{\partial \boldsymbol{\theta}},
\end{aligned} \tag{D.2}$$

with

$$A_t^* := \frac{\partial \bar{p}_t}{\partial f_t} + \frac{\partial \log \bar{g}'_t}{\partial f_t} = \nabla_t, \quad \bar{\mathbf{f}}_1^{0:1} = (\bar{f}_1, \partial \bar{f}_1 / \partial \boldsymbol{\theta}),$$

and

$$\frac{\partial f_t}{\partial \boldsymbol{\theta}} = \begin{bmatrix} \frac{\partial f_t}{\partial \omega} & \frac{\partial f_t}{\partial \alpha} & \frac{\partial f_t}{\partial \beta} & \frac{\partial f_t}{\partial \lambda} \end{bmatrix}^\top, \quad \frac{\partial \bar{p}_t}{\partial \boldsymbol{\theta}} := \begin{bmatrix} 0 & 0 & 0 & \frac{\partial \bar{p}_t}{\partial \lambda} \end{bmatrix}^\top,$$

and  $\mathbf{f}_t^{(0:i)}$  denoting  $f_t$  and its first  $i$  derivatives with respect to  $\boldsymbol{\theta}$ . The second derivative of the log-likelihood function is given by

$$\begin{aligned}
\ell_T''(\boldsymbol{\theta}, \bar{\mathbf{f}}_1^{0:2}) &= \frac{\partial^2 \ell(\boldsymbol{\theta}, \bar{\mathbf{f}}_1)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \\
&= \frac{1}{T} \sum_{t=1}^T \left( \frac{\partial^2 f_t}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \cdot A_t^* + \frac{\partial f_t}{\partial \boldsymbol{\theta}} \cdot \frac{\partial A_t^*}{\partial f_t} \cdot \frac{\partial f_t}{\partial \boldsymbol{\theta}^\top} + \frac{\partial f_t}{\partial \boldsymbol{\theta}} \cdot \frac{\partial A_t^*}{\partial \boldsymbol{\theta}^\top} + \frac{\partial^2 \bar{p}_t}{\partial \boldsymbol{\theta} \partial f_t} \frac{\partial f_t}{\partial \boldsymbol{\theta}^\top} + \frac{\partial^2 \bar{p}_t}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \right) \\
&= \frac{1}{T} \sum_{t=1}^T \left( \frac{\partial^2 f_t}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \cdot A_t^* + \frac{\partial f_t}{\partial \boldsymbol{\theta}} \frac{\partial f_t}{\partial \boldsymbol{\theta}^\top} \cdot B_t^* + \frac{\partial f_t}{\partial \boldsymbol{\theta}} (C_t^*)^\top + C_t^* \frac{\partial f_t}{\partial \boldsymbol{\theta}^\top} + \frac{\partial^2 \bar{p}_t}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \right), \\
&= \frac{1}{T} \sum_{t=1}^T \left( \frac{\partial^2 f_t}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \cdot \nabla_t + \frac{\partial f_t}{\partial \boldsymbol{\theta}} \frac{\partial f_t}{\partial \boldsymbol{\theta}^\top} \frac{\partial \nabla_t}{\partial f_t} + \frac{\partial f_t}{\partial \boldsymbol{\theta}} \frac{\partial \nabla_t}{\partial \boldsymbol{\theta}^\top} + \frac{\partial \nabla_t}{\partial \boldsymbol{\theta}} \frac{\partial f_t}{\partial \boldsymbol{\theta}^\top} + \frac{\partial^2 \bar{p}_t}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \right),
\end{aligned} \tag{D.3}$$

where

$$\begin{aligned}
\bar{\mathbf{f}}_1^{0:2} &= \left( \bar{f}_1, \partial \bar{f}_1 / \partial \boldsymbol{\theta}, \partial^2 \bar{f}_1 / \partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top \right), \\
B_t^* &= \frac{\partial^2 \bar{p}_t}{\partial f_t^2} + \frac{\partial^2 \log \bar{g}_t'}{\partial f_t^2} = \frac{\partial \nabla_t}{\partial f_t}, \\
C_t^* &= \left[ 0 \ 0 \ 0 \ \frac{\partial^2 \bar{p}_t}{\partial f_t \partial \lambda} \right]^\top = \left[ 0 \ 0 \ 0 \ \frac{\partial \nabla_t}{\partial \lambda} \right]^\top = \frac{\partial A_t^*}{\partial \boldsymbol{\theta}^\top}, \\
\frac{\partial^2 f_t}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} &= \begin{bmatrix} \frac{\partial^2 f_t}{\partial \omega^2} & \frac{\partial^2 f_t}{\partial \omega \partial \alpha} & \frac{\partial^2 f_t}{\partial \omega \partial \beta} & \frac{\partial^2 f_t}{\partial \omega \partial \lambda} \\ \frac{\partial^2 f_t}{\partial \alpha \partial \omega} & \frac{\partial^2 f_t}{\partial \alpha^2} & \frac{\partial^2 f_t}{\partial \alpha \partial \beta} & \frac{\partial^2 f_t}{\partial \alpha \partial \lambda} \\ \frac{\partial^2 f_t}{\partial \beta \partial \omega} & \frac{\partial^2 f_t}{\partial \beta \partial \alpha} & \frac{\partial^2 f_t}{\partial \beta^2} & \frac{\partial^2 f_t}{\partial \beta \partial \lambda} \\ \frac{\partial^2 f_t}{\partial \lambda \partial \omega} & \frac{\partial^2 f_t}{\partial \lambda \partial \alpha} & \frac{\partial^2 f_t}{\partial \lambda \partial \beta} & \frac{\partial^2 f_t}{\partial \lambda^2} \end{bmatrix}, \\
\frac{\partial^2 \bar{p}_t}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} &= \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{\partial^2 \bar{p}_t}{\partial \lambda^2} \end{bmatrix}.
\end{aligned}$$

## D.2 Expressions for the derivative processes of $f_t$

We have  $\boldsymbol{\theta} = (\omega, \alpha, \beta, \lambda) \in \Theta$  and write  $\partial s(f_t, v_t; \lambda) / \partial \boldsymbol{\theta}_i$  as the derivative of the scaled score w.r.t.  $\lambda$  only, not accounting for the dependence of  $f_t$  on  $\boldsymbol{\theta}$ . Differentiating the transition equation of the score-driven model, we obtain

$$\begin{aligned}
\frac{\partial f_{t+1}}{\partial \boldsymbol{\theta}_i} &= \frac{\partial \omega}{\partial \boldsymbol{\theta}_i} + \frac{\partial \alpha}{\partial \boldsymbol{\theta}_i} s_t + \alpha \frac{\partial s_t}{\partial f_t} \frac{\partial f_t}{\partial \boldsymbol{\theta}_i} + \alpha \frac{\partial s_t}{\partial \boldsymbol{\theta}_i} + \frac{\partial \beta}{\partial \boldsymbol{\theta}_i} f_t + \beta \frac{\partial f_t}{\partial \boldsymbol{\theta}_i}, \\
&= \mathbf{A}_{i,t}^{(1)} + \frac{\partial f_t}{\partial \boldsymbol{\theta}_i} \mathbf{B}_t,
\end{aligned} \tag{D.4}$$

with

$$\begin{aligned}
\mathbf{A}_t^{(1)} &= \mathbf{A}_t^{(1)}(f_t, \boldsymbol{\theta}) = (\mathbf{A}_{1,t}^{(1)}(f_t, \boldsymbol{\theta}), \dots, \mathbf{A}_{4,t}^{(1)}(f_t, \boldsymbol{\theta}))^\top = \frac{\partial \omega}{\partial \boldsymbol{\theta}} + \frac{\partial \alpha}{\partial \boldsymbol{\theta}} s_t + \alpha \frac{\partial s_t}{\partial \boldsymbol{\theta}} + \frac{\partial \beta}{\partial \boldsymbol{\theta}} f_t, \\
\mathbf{B}_t &= \mathbf{B}_t(f_t, \boldsymbol{\theta}) = \alpha \frac{\partial s_t}{\partial f_t} + \beta.
\end{aligned}$$



For the second derivative process, we obtain a recursion

$$\begin{aligned}\frac{\partial^2 f_{t+1}}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} &= \frac{\partial \mathbf{A}_t^{(1)}}{\partial \boldsymbol{\theta}^\top} + \frac{\partial \mathbf{A}_t^{(1)}}{\partial f_t} \frac{\partial f_t}{\partial \boldsymbol{\theta}^\top} + \frac{\partial f_t}{\partial \boldsymbol{\theta}} \frac{\partial \mathbf{B}_t}{\partial \boldsymbol{\theta}^\top} + \frac{\partial f_t}{\partial \boldsymbol{\theta}} \frac{\partial \mathbf{B}_t}{\partial f_t} \frac{\partial f_t}{\partial \boldsymbol{\theta}^\top} + \frac{\partial^2 f_t}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \mathbf{B}_t \\ &= \mathbf{A}_t^{(2)} + \frac{\partial^2 f_t}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \mathbf{B}_t,\end{aligned}\quad (\text{D.5})$$

with

$$\begin{aligned}\mathbf{A}_t^{(2)} &= \frac{\partial \mathbf{A}_t^{(1)}}{\partial \boldsymbol{\theta}^\top} + \frac{\partial \mathbf{A}_t^{(1)}}{\partial f_t} \frac{\partial f_t}{\partial \boldsymbol{\theta}^\top} + \frac{\partial f_t}{\partial \boldsymbol{\theta}} \frac{\partial \mathbf{B}_t}{\partial \boldsymbol{\theta}^\top} + \frac{\partial \mathbf{B}_t}{\partial f_t} \frac{\partial f_t}{\partial \boldsymbol{\theta}} \frac{\partial f_t}{\partial \boldsymbol{\theta}^\top} \\ &= \left( \frac{\partial \alpha}{\partial \boldsymbol{\theta}} \frac{\partial s_t}{\partial \boldsymbol{\theta}^\top} + \frac{\partial s_t}{\partial \boldsymbol{\theta}} \frac{\partial \alpha}{\partial \boldsymbol{\theta}^\top} + \alpha \frac{\partial^2 s_t}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \right) + \left( \frac{\partial \alpha}{\partial \boldsymbol{\theta}} \frac{\partial s_t}{\partial f_t} + \alpha \frac{\partial^2 s_t}{\partial \boldsymbol{\theta} \partial f_t} + \frac{\partial \beta}{\partial \boldsymbol{\theta}} \right) \frac{\partial f_t}{\partial \boldsymbol{\theta}^\top} \\ &\quad + \frac{\partial f_t}{\partial \boldsymbol{\theta}} \left( \frac{\partial s_t}{\partial f_t} \frac{\partial \alpha}{\partial \boldsymbol{\theta}^\top} + \alpha \frac{\partial^2 s_t}{\partial f_t \partial \boldsymbol{\theta}^\top} + \frac{\partial \beta}{\partial \boldsymbol{\theta}^\top} \right) + \alpha \frac{\partial^2 s_t}{\partial f_t^2} \frac{\partial f_t}{\partial \boldsymbol{\theta}} \frac{\partial f_t}{\partial \boldsymbol{\theta}^\top}.\end{aligned}\quad (\text{D.6})$$

## E Further Technical Lemmas and Proofs

This appendix contains a number of more technical results, some of which are more generic than the results in the main paper. These technical results are used to prove the propositions from the main paper.

We start with two more generic propositions to prove Propositions 3.1 and 3.3 from the main text. The propositions below encompass the score-driven model and are written for the case of general random sequences  $\{x_t(\boldsymbol{\theta}, \bar{x})\}_{t \in \mathbb{N}}$  taking values in  $\mathcal{X} \subseteq \mathbb{R}$ , where  $x_t(\boldsymbol{\theta}, \bar{x})$  is generated by a stochastic recurrence equation of the form

$$x_{t+1}(\boldsymbol{\theta}, \bar{x}) = \phi(x_t(\boldsymbol{\theta}, \bar{x}), v_t, \boldsymbol{\theta}), \quad (\text{E.1})$$

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

$$r_t^k(\boldsymbol{\theta}) := \sup_{(x, x') \in \mathcal{X}^* \times \mathcal{X}^* : x \neq x'} \frac{|\phi(x, v_t, \boldsymbol{\theta}) - \phi(x', v_t, \boldsymbol{\theta})|^k}{|x - x'|^k}, \quad k \geq 0.$$

Moreover, for random sequences  $\{x_{1,t}\}_{t \in \mathbb{Z}}$  and  $\{x_{2,t}\}_{t \in \mathbb{Z}}$ , we say that  $x_{1,t}$  converges exponentially fast almost surely (e.a.s.) to  $x_{2,t}$  if there exists a constant  $c > 1$  such that  $c^t \|x_{1,t} - x_{2,t}\| \xrightarrow{\text{a.s.}} 0$ ; see also Straumann and Mikosch (2006) (hereafter referred to as SM06).

**Proposition TA.5.** *For every  $\boldsymbol{\theta} \in \Theta$ , let  $\{v_t\}_{t \in \mathbb{Z}}$  be a strictly stationary and ergodic (SE) sequence and assume  $\exists \bar{x} \in \mathcal{X}$  such that*

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

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

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

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

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

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

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

then  $\sup_t \mathbb{E}|x_t(\boldsymbol{\theta}, \bar{x})|^n < \infty$  and  $\mathbb{E}|x_t(\boldsymbol{\theta})|^{n-\delta} < \infty \forall \boldsymbol{\theta} \in \Theta$  and arbitrarily small  $\delta > 0$ .

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

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

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

then  $\sup_t \mathbb{E}|x_t(\boldsymbol{\theta}, \bar{x})|^n < \infty$  and  $\mathbb{E}|x_t(\boldsymbol{\theta})|^{n-\delta} < \infty \forall \boldsymbol{\theta} \in \Theta$  and every  $n > 0$  and arbitrarily small  $\delta > 0$ .

*Proof of Proposition TA.5. Step 1, SE:* The assumption that  $\{v_t\}_{t \in \mathbb{Z}}$  is SE  $\forall \boldsymbol{\theta} \in \Theta$  together with the continuity of  $\phi$  on  $\mathcal{X} \times \mathcal{V} \times \Theta$  (and resulting measurability w.r.t. the Borel  $\sigma$ -algebra) implies that  $\{\phi_t := \phi(\cdot, v_t, \boldsymbol{\theta})\}_{t \in \mathbb{Z}}$  is SE for every  $\boldsymbol{\theta} \in \Theta$  by Krengel (1985, Proposition 4.3). Condition C1 in Bougerol (1993, Theorem 3.1) is immediately implied by assumption (i) for every  $\boldsymbol{\theta} \in \Theta$ . Condition C2 in Bougerol (1993, Theorem 3.1) is implied, for every  $\boldsymbol{\theta} \in \Theta$ , by condition (ii) since for every  $\boldsymbol{\theta} \in \Theta$ ,

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

Also due to the stationarity of  $\{v_t\}$  we have  $\mathbb{E} \log r_t^1(\boldsymbol{\theta}) = \mathbb{E} \log r_1^1(\boldsymbol{\theta})$ . As a result, for every  $\boldsymbol{\theta} \in \Theta$ ,  $\{x_t(\boldsymbol{\theta}, \bar{x})\}_{t \in \mathbb{N}}$  converges to an SE solution  $\{x_t(v_{\boldsymbol{\theta}}^{t-1}, \boldsymbol{\theta})\}_{t \in \mathbb{Z}}$ . Uniqueness and e.a.s. convergence are obtained by Straumann and Mikosch (2006, Theorem 2.8).

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

$$\sup_t \|x_t(\boldsymbol{\theta}, \bar{x}) - \bar{x}_{\boldsymbol{\theta}}\|_n \tag{E.2}$$

$$\begin{aligned} &= \sup_t \|\phi(x_{t-1}(\boldsymbol{\theta}, \bar{x}), v_{t-1}(\boldsymbol{\theta}), \boldsymbol{\theta}) - \phi(\bar{x}, \bar{v}, \boldsymbol{\theta})\|_n \\ &\leq \sup_t \|\phi(x_{t-1}(\boldsymbol{\theta}, \bar{x}), v_{t-1}(\boldsymbol{\theta}), \boldsymbol{\theta}) - \phi(\bar{x}, v_{t-1}(\boldsymbol{\theta}), \boldsymbol{\theta})\|_n \\ &\quad + \sup_t \|\phi(\bar{x}, v_{t-1}(\boldsymbol{\theta}), \boldsymbol{\theta})\|_n + |\phi(\bar{x}, \bar{v}, \boldsymbol{\theta})| \\ &\leq \sup_t \left( \mathbb{E}|x_{t-1}(\boldsymbol{\theta}, \bar{x}) - \bar{x}|^n \right. \\ &\quad \left. \times \frac{|\phi(x_{t-1}(\boldsymbol{\theta}, \bar{x}), v_{t-1}(\boldsymbol{\theta}), \boldsymbol{\theta}) - \phi(\bar{x}, v_{t-1}(\boldsymbol{\theta}), \boldsymbol{\theta})|^n}{|x_{t-1}(\boldsymbol{\theta}, \bar{x}) - \bar{x}|^n} \right)^{1/n} \\ &\quad + \sup_t \|\phi(\bar{x}, v_{t-1}(\boldsymbol{\theta}), \boldsymbol{\theta})\|_n + |\bar{x}_{\boldsymbol{\theta}}| \end{aligned}$$

$$\begin{aligned} &\leq \sup_t \left( \mathbb{E}|x_{t-1}(\boldsymbol{\theta}, \bar{x}) - \bar{x}_{\boldsymbol{\theta}} + \bar{x}_{\boldsymbol{\theta}} - \bar{x}|^n \times r_t^n(\boldsymbol{\theta}) \right)^{1/n} \\ &\quad + \sup_t \|\phi(\bar{x}, v_{t-1}(\boldsymbol{\theta}), \boldsymbol{\theta})\|_n + |\bar{x}_{\boldsymbol{\theta}}| \\ &\leq \sup_t \|x_{t-1}(\boldsymbol{\theta}, \bar{x}) - \bar{x}_{\boldsymbol{\theta}}\|_n \cdot (\mathbb{E}r_t^n(\boldsymbol{\theta}))^{1/n} \tag{E.3} \\ &\quad + \sup_t \|\phi(\bar{x}, v_{t-1}(\boldsymbol{\theta}), \boldsymbol{\theta})\|_n + |\bar{x}_{\boldsymbol{\theta}} - \bar{x}| (\mathbb{E}r_t^n(\boldsymbol{\theta}))^{1/n} + |\bar{x}_{\boldsymbol{\theta}}| \end{aligned}$$

$$\begin{aligned} &\leq \bar{c}_n \cdot \sup_t \|x_{t-1}(\boldsymbol{\theta}, \bar{x}) - \bar{x}_{\boldsymbol{\theta}}\|_n + \|\phi(\bar{x}, v_1(\boldsymbol{\theta}), \boldsymbol{\theta})\|_n + (1 + \bar{c}_n)|\bar{x}_{\boldsymbol{\theta}}| + \bar{c}_n |\bar{x}|, \tag{E.4} \end{aligned}$$

where  $\bar{c}_n := (\mathbb{E}r_t^n(\boldsymbol{\theta}))^{1/n} = (\mathbb{E}r_1^n(\boldsymbol{\theta}))^{1/n} < 1$  by assumption (iv.a) and the stationarity of  $\{v_t\}$ . The step to (E.3) can be made due to the independence assumed in condition (v.a), and the step to (E.4) again uses the stationarity of  $\{v_t\}$ . We can now unfold the recursion (E.2)–(E.4) as

$$\begin{aligned} \sup_t \|x_t(\boldsymbol{\theta}, \bar{x}) - \bar{x}_\boldsymbol{\theta}\|_n &\leq \\ &(\bar{c}_n)^{t-1} |\bar{x} - \bar{x}_\boldsymbol{\theta}| + \sum_{j=0}^{t-2} (\bar{c}_n)^j \left( \|\phi(\bar{x}, v_1(\boldsymbol{\theta}), \boldsymbol{\theta})\|_n + (1 + \bar{c}_n) |\bar{x}_\boldsymbol{\theta}| + \bar{c}_n |\bar{x}| \right) \leq \\ &|\bar{x} - \bar{x}_\boldsymbol{\theta}| + \sum_{j=0}^{t-2} (\bar{c}_n)^j \left( \|\phi(\bar{x}, v_1(\boldsymbol{\theta}), \boldsymbol{\theta})\|_n + (1 + \bar{c}_n) |\bar{x}_\boldsymbol{\theta}| + \bar{c}_n |\bar{x}| \right) \leq \\ &|\bar{x} - \bar{x}_\boldsymbol{\theta}| + \frac{\|\phi(\bar{x}, v_1(\boldsymbol{\theta}), \boldsymbol{\theta})\|_n + 2|\bar{x}_\boldsymbol{\theta}| + |\bar{x}|}{1 - \bar{c}_n} < \infty \quad \forall \boldsymbol{\theta} \in \Theta. \end{aligned}$$

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

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

As a result, unfolding the recursion renders  $\sup_t \|x_t(\boldsymbol{\theta}, \bar{x})\|_n < \infty$  by the same argument as above. For  $0 < n < 1$  the function  $\|\cdot\|_n$  is only a pseudo-norm as it is not sub-additive. However, the proof still follows by working instead with the metric  $\|\cdot\|_n^* := (\|\cdot\|_n)^n$  which is sub-additive; see the  $C_n$  inequality in Loève (1977).

The moments of the limit process  $x_t(\boldsymbol{\theta})$  can finally be obtained by noting that the limit of a sequence satisfying  $\|x_t(\boldsymbol{\theta}, \bar{x})\|_m < \infty \forall t$ , that converges almost surely  $|x_t(\boldsymbol{\theta}, \bar{x}) - x_t(\boldsymbol{\theta})| \xrightarrow{a.s.} 0$ , and is uniformly integrable  $\lim_{M \rightarrow \infty} \sup_t \mathbb{E}[\|x_t(\boldsymbol{\theta}, \bar{x})\|_m \cdot \mathbf{1}_{(|x_t(\boldsymbol{\theta}, \bar{x})| > M)}] = 0$  satisfies, by Theorem 2.17 in Jiang (2010),  $\mathbb{E}\|x_t(\boldsymbol{\theta})\|_m < \infty$ . We already proved finite moments for the initialized sequence  $x_t(\boldsymbol{\theta}, \bar{x})$ , and the almost sure convergence is implied by the exponentially fast almost sure convergence. The uniform integrability condition for  $m = n - \delta$  for arbitrarily small  $\delta > 0$  follows by adapting Theorem 12.10 in Davidson (1994), since

$$\begin{aligned} \infty &> \left( \sup_t \|x_t(\boldsymbol{\theta}, \bar{x})\|_n \right)^n = \sup_t \mathbb{E}[x_t(\boldsymbol{\theta}, \bar{x})]^n \\ &\geq \sup_t \mathbb{E}[x_t(\boldsymbol{\theta}, \bar{x})^n \mathbf{1}_{(|x_t(\boldsymbol{\theta}, \bar{x})| > M)}] \\ &\geq \sup_t \mathbb{E}\left[x_t(\boldsymbol{\theta}, \bar{x})^m \mathbf{1}_{(|x_t(\boldsymbol{\theta}, \bar{x})| > M)} x_t(\boldsymbol{\theta}, \bar{x})^\delta \mathbf{1}_{(|x_t(\boldsymbol{\theta}, \bar{x})| > M)}\right] \\ &\geq M^\delta \sup_t \mathbb{E}[x_t(\boldsymbol{\theta}, \bar{x})^m \mathbf{1}_{(|x_t(\boldsymbol{\theta}, \bar{x})| > M)}]. \end{aligned}$$

Hence, given the boundedness of the left-hand side and letting  $M \rightarrow \infty$ , we obtain for the right-hand side that  $\lim_{M \rightarrow \infty} \sup_t \mathbb{E}[x_t(\boldsymbol{\theta}, \bar{x})^m \mathbf{1}_{(|x_t(\boldsymbol{\theta}, \bar{x})| > M)}] = 0$ , which establishes the result.  $\blacksquare$

Proposition TA.5 not only establishes the convergence to a unique SE solution, but also establishes the existence of unconditional moments. The latter property is key to proving the consistency and asymptotic normality of the MLE in Section 4 of the paper. To establish convergence to an

SE solution, condition (ii) requires the stochastic recurrence equation to be contracting on average. For the subsequent existence of moments, the contraction condition (iv.a), together with the moment bound in (iii.a), and the independence assumption (v.a), are sufficient. Alternatively, if by condition (iii.b)  $\phi$  is uniformly bounded in  $v$ , then a deterministic contraction condition (iv.b) only needs to hold on the uniform bound. In that case, we also obtain the stronger result that the moment bound holds for any  $n > 0$ . Note that conditions (i)–(ii) are implied by (iii.a)–(v.a). Remark TA.6 shows that condition (v.a) is automatically satisfied if  $v_t$  is an innovation sequence.

**Remark TA.6.** If  $v_t \perp x_t(\boldsymbol{\theta}, \bar{x})$  then (v.a) in Proposition TA.5 holds.

The condition that  $v_t$  is an innovation sequence is typically more intuitive. We keep the independence assumption (v.a) in Proposition TA.5, however, because in some of our models the supremum Lipschitz constant is independent of the random  $v_t$ . In such cases, the independence is easily satisfied, even in cases where  $v_t$  is not an innovation sequence.

Following SM06, we also note that conditions (i) and (ii) in Proposition TA.5 provide us with an almost sure representation of  $x_t(\boldsymbol{\theta}, \bar{x})$  as a measurable function of  $\{v_s(\boldsymbol{\theta})\}_{s \leq t-1}$ . Let  $\circ$  denote the composition of maps, e.g.,

$$\phi(\cdot, v_{t-1}(\boldsymbol{\theta}), \boldsymbol{\theta}) \circ \phi(\cdot, v_{t-2}(\boldsymbol{\theta}), \boldsymbol{\theta}) = \phi\left(\phi(\cdot, v_{t-2}(\boldsymbol{\theta}), \boldsymbol{\theta}), v_{t-1}(\boldsymbol{\theta}), \boldsymbol{\theta}\right).$$

We have the following Lemma.

**Remark TA.7.** Let conditions (i) and (ii) of Proposition TA.5 hold. Then  $x_t(\boldsymbol{\theta})$  admits the following a.s. representation for every  $\boldsymbol{\theta} \in \Theta$

$$x_t(\boldsymbol{\theta}) = \lim_{r \rightarrow \infty} \phi(\cdot, v_{t-1}(\boldsymbol{\theta}), \boldsymbol{\theta}) \circ \phi(\cdot, v_{t-2}(\boldsymbol{\theta}), \boldsymbol{\theta}) \circ \dots \circ \phi(\cdot, v_{t-r}(\boldsymbol{\theta}), \boldsymbol{\theta}),$$

and  $x_t(\boldsymbol{\theta})$  is measurable with respect to the  $\sigma$ -algebra generated by  $\{v_s(\boldsymbol{\theta})\}_{s \leq t-1}$ .

Proposition TA.8 deals with sequences  $\{x_t(\boldsymbol{\theta}, \bar{x})\}_{t \in \mathbb{N}}$  that, for a given initialization  $\bar{x} \in \mathcal{X}$ , are generated by

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

where  $\phi : \mathcal{X} \times \mathcal{V} \times \Theta \rightarrow \mathcal{X}$  is continuous. We have the following proposition.

**Proposition TA.8.** Let  $\Theta$  be compact,  $\{v_t\}_{t \in \mathbb{Z}}$  be stationary and ergodic (SE) and assume there exists an  $\bar{x} \in \mathcal{X}$ , such that

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

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

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

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

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

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

or

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

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

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

then  $\sup_t \mathbb{E} \sup_{\theta \in \Theta} |x_t(\theta, \bar{x})|^n < \infty$  and  $\mathbb{E} \sup_{\theta \in \Theta} |x_t(\theta)|^{n-\delta} < \infty$  and arbitrarily small  $\delta > 0$ .

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

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

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

then  $\sup_t \mathbb{E} \sup_{\theta \in \Theta} |x_t(\theta, \bar{x})|^n < \infty$  and  $\mathbb{E} \sup_{\theta \in \Theta} |x_t(\theta)|^{n-\delta} < \infty$  for every  $n > 0$  and arbitrarily small  $\delta > 0$ .

The contraction condition (iv.a) in Proposition TA.8 is stricter than condition (iv.b). Rather than only requiring the contraction property to hold in expectation, condition (iv.a) holds for all  $v \in \mathcal{V}$ .

Again, we note that conditions (i) and (ii) in Proposition TA.8 provide us with an almost sure representation of  $x_t(\theta)$  in terms of .

**Remark TA.9.** Let conditions (i) and (ii) of Proposition TA.8 hold. Then  $x_t(\theta)$  admits the following a.s. representation for every  $\theta \in \Theta$

$$x_t(\theta) = \lim_{r \rightarrow \infty} \phi(\cdot, v_{t-1}, \theta) \circ \phi(\cdot, v_{t-2}, \theta) \circ \dots \circ \phi(\cdot, v_{t-r}, \theta)$$

and  $x_t(\theta)$  is measurable with respect to the  $\sigma$ -algebra generated by .

Inspection of the proof of Proposition TA.8 reveals that the moment bounds hold even if the innovations are non-stationary.

**Remark TA.10.** Let  $\{v_t\}$  be non-stationary and suppose that conditions (iii.b), (iv.b) and (v.b) hold uniformly in  $t$

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

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

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

Then  $\sup_t \mathbb{E} \sup_{\theta \in \Theta} |x_t(\theta, \bar{x})|^n < \infty$  and  $\sup_t \mathbb{E} \sup_{\theta \in \Theta} |x_t(\theta)|^{n-\delta} < \infty$  for every  $n > 0$  and arbitrarily small  $\delta > 0$ .

*Proof of Proposition TA.8. Step 0, additional notation:* Following Straumann and Mikosch (2006, Proposition 3.12), the uniform convergence of the process  $\sup_{\theta \in \Theta} |x_t(\theta, \bar{x}) - x_t(\theta)| \xrightarrow{e.a.s.} 0$  is obtained by appealing to Bougerol (1993, Theorem 3.1) using sequences of random functions  $\{x_t(\cdot, \bar{x})\}_{t \in \mathbb{N}}$  rather than sequences of real numbers. This change is subtle in the notation, but important. We refer to SM06 for details.

The elements  $x_t(\cdot, \bar{x})$  are random functions that take values in the separable Banach space  $\mathcal{X}_\Theta \subseteq (\mathbb{C}(\Theta, \mathcal{X}), \|\cdot\|^\Theta)$ , where  $\|x_t(\cdot)\|_n^\Theta \equiv (\mathbb{E} \sup_{\theta \in \Theta} |x_t(\theta)|^n)^{1/n}$  and  $\|x_t(\cdot)\|^\Theta = \|x_t(\cdot)\|_1^\Theta$ . The functions  $x_t(\cdot, \bar{x})$  are generated by

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

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

3.12). Note the subtle but important difference between  $\phi^*(\cdot, v_t, \cdot) : \mathbb{C}(\Theta) \times \Theta \rightarrow \mathbb{C}(\Theta)$  and  $\phi(\cdot, v_t, \cdot) : \mathcal{X} \times \Theta \rightarrow \mathcal{X}$  as alluded to earlier.

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

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

$$\begin{aligned}
& \mathbb{E} \log \sup_{\|\bar{x}_\Theta - \bar{x}'_\Theta\|^\Theta > 0} \frac{\|\phi^*(\bar{x}_\Theta(\cdot), v_t, \cdot) - \phi^*(\bar{x}'_\Theta(\cdot), v_t, \cdot)\|^\Theta}{\|\bar{x}_\Theta(\cdot) - \bar{x}'_\Theta(\cdot)\|^\Theta} = \\
& \mathbb{E} \log \sup_{\|\bar{x}_\Theta - \bar{x}'_\Theta\|^\Theta > 0} \frac{\sup_{\theta \in \Theta} |\phi(\bar{x}_\Theta(\theta), v_t, \theta) - \phi(\bar{x}'_\Theta(\theta), v_t, \theta)|}{\|\bar{x}_\Theta(\cdot) - \bar{x}'_\Theta(\cdot)\|^\Theta} = \\
& \mathbb{E} \log \sup_{\|\bar{x}_\Theta - \bar{x}'_\Theta\|^\Theta > 0} \sup_{\theta \in \Theta} \frac{|\phi(\bar{x}_\Theta(\theta), v_t, \theta) - \phi(\bar{x}'_\Theta(\theta), v_t, \theta)|}{\|\bar{x}_\Theta(\cdot) - \bar{x}'_\Theta(\cdot)\|^\Theta} = \\
& \mathbb{E} \log \sup_{\|\bar{x}_\Theta - \bar{x}'_\Theta\|^\Theta > 0} \sup_{\theta \in \Theta | \bar{x}_\Theta(\theta) \neq \bar{x}'_\Theta(\theta)} \frac{|\phi(\bar{x}_\Theta(\theta), v_t, \theta) - \phi(\bar{x}'_\Theta(\theta), v_t, \theta)|}{\|\bar{x}_\Theta(\cdot) - \bar{x}'_\Theta(\cdot)\|^\Theta} \leq \\
& \mathbb{E} \log \sup_{\|\bar{x}_\Theta - \bar{x}'_\Theta\|^\Theta > 0} \sup_{\theta \in \Theta | \bar{x}_\Theta(\theta) \neq \bar{x}'_\Theta(\theta)} \frac{|\phi(\bar{x}_\Theta(\theta), v_t, \theta) - \phi(\bar{x}'_\Theta(\theta), v_t, \theta)|}{|\bar{x}_\Theta(\theta) - \bar{x}'_\Theta(\theta)|} \times \\
& \frac{|\bar{x}_\Theta(\theta) - \bar{x}'_\Theta(\theta)|}{\|\bar{x}_\Theta(\cdot) - \bar{x}'_\Theta(\cdot)\|^\Theta} \leq \\
& \mathbb{E} \log \left( \sup_{\|\bar{x}_\Theta - \bar{x}'_\Theta\|^\Theta > 0} \sup_{\theta \in \Theta | \bar{x}_\Theta(\theta) \neq \bar{x}'_\Theta(\theta)} \frac{|\phi(\bar{x}_\Theta(\theta), v_t, \theta) - \phi(\bar{x}'_\Theta(\theta), v_t, \theta)|}{|\bar{x}_\Theta(\theta) - \bar{x}'_\Theta(\theta)|} \right) \times \\
& \left( \sup_{\|\bar{x}_\Theta - \bar{x}'_\Theta\|^\Theta > 0} \sup_{\theta \in \Theta} \frac{|\bar{x}_\Theta(\theta) - \bar{x}'_\Theta(\theta)|}{\|\bar{x}_\Theta(\cdot) - \bar{x}'_\Theta(\cdot)\|^\Theta} \right) \leq \\
& \mathbb{E} \log \sup_{\|\bar{x}_\Theta - \bar{x}'_\Theta\|^\Theta > 0} \sup_{\theta \in \Theta} \sup_{x \neq x'} \frac{|\phi(x, v_t, \theta) - \phi(x', v_t, \theta)|}{|x - x'|} = \\
& \mathbb{E} \log \sup_{\theta \in \Theta} \sup_{x \neq x'} \frac{|\phi(x, v_t, \theta) - \phi(x', v_t, \theta)|}{|x - x'|} = \mathbb{E} \log \sup_{\theta \in \Theta} r_t^1(\theta) = \\
& \mathbb{E} \log \sup_{\theta \in \Theta} r_1^1(\theta) < 0.
\end{aligned}$$

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

*Step 2, moment bounds:* We use a similar argument as in the proof of Proposition TA.5. First consider  $n \geq 1$  and note that  $\sup_t \mathbb{E} \sup_{\theta \in \Theta} |x_t(\theta, \bar{x})|^n < \infty$  if and only if  $\sup_t \|x_t(\theta, \bar{x})\|_n^\Theta < \infty$ . Further,  $\|x_t(\cdot, \bar{x}) - \bar{x}_\Theta\|_n^\Theta < \infty$  implies  $\|x_t(\cdot, \bar{x})\|_n^\Theta < \infty$  for any  $\bar{x}_\Theta \in \mathcal{X}_\Theta \subseteq \mathbb{C}(\Theta)$ , since continuity on the compact  $\Theta$  implies  $\sup_{\theta \in \Theta} |\bar{x}_\Theta(\theta)| < \infty$ . For a pair  $(\bar{x}, \bar{v}) \in \mathcal{X} \times \mathcal{V}$ , let  $\bar{x}_\Theta(\cdot) = \phi(\bar{x}, \bar{v}, \cdot) \in \mathbb{C}(\Theta)$ . By compactness of  $\Theta$  and continuity of  $\bar{x}_\Theta$  we immediately have  $\bar{x}_\Theta := \|\bar{x}_\Theta(\cdot)\|_n^\Theta < \infty$ . Also  $\bar{\phi} := \sup_t \|\phi(\bar{x}, v_t, \cdot)\|_n^\Theta < \infty$  by condition (iii.a). Using condition (iv.a), define  $\bar{c} < 1$  such that

$\sup_{\theta \in \Theta} |\phi(x, v, \theta) - \phi(x', v, \theta)| \leq \bar{c} |x - x'|$  for all  $(x, x', v)$ . We now have

$$\begin{aligned}
& \sup_t \|x_t(\cdot, \bar{x}) - \bar{x}_\Theta(\cdot)\|_n^\Theta \\
& \leq \sup_t \|\phi^*(x_{t-1}(\cdot, \bar{x}), v_t, \cdot) - \phi(\bar{x}, v_t, \cdot)\|_n^\Theta \\
& \quad + \sup_t \|\phi(\bar{x}, v_t, \cdot)\|_n^\Theta + \sup_t \|\bar{x}_\Theta(\cdot)\|_n^\Theta \\
& \leq \sup_t \|\phi^*(x_{t-1}(\cdot, \bar{x}), v_t, \cdot) - \phi(\bar{x}, v_t, \cdot)\|_n^\Theta + \bar{\phi} + \bar{x}_\Theta \\
& \leq \bar{c} \cdot \sup_t \|x_{t-1}(\cdot, \bar{x}) - \bar{x}\|_n^\Theta + \bar{\phi} + \bar{x}_\Theta \\
& \leq \bar{c} \cdot \sup_t \|x_{t-1}(\cdot, \bar{x}) - \bar{x}_\Theta\|_n^\Theta + \bar{\phi} + \bar{c} \bar{x} + (1 + \bar{c}) \bar{x}_\Theta \\
& \leq \left( \bar{\phi} + \bar{c} \bar{x} + (1 + \bar{c}) \bar{x}_\Theta \right) \sum_{j=0}^{t-2} \bar{c}^j + \bar{c}^{t-1} \sup_t \|\bar{x} - \bar{x}_\Theta(\cdot)\|_n^\Theta \\
& \leq \frac{(\bar{\phi} + \bar{c} \bar{x} + (1 + \bar{c}) \bar{x}_\Theta)}{1 - \bar{c}} + |\bar{x}| + \bar{x}_\Theta < \infty,
\end{aligned}$$

using  $\bar{c} < 1$ .

To use conditions (iii.b)–(v.b) instead of conditions (iii.a)–(v.a), we first (re)define  $\bar{c}$  as  $\bar{c}_n = (\sup_{\theta \in \Theta} r_1^1(\theta))^{1/n} = (\sup_{\theta \in \Theta} r_t^1(\theta))^{1/n}$ . Also using the independence assumed in condition (v.b), we have

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

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

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

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

with  $\|\bar{\phi}(\bar{x}, \cdot)\|_n^\Theta < \infty$  by (iii.c) and  $\bar{c} < 1$  by condition (iv.c). As a result, unfolding the recursion establishes  $\sup_t \|x_t(\cdot, \bar{x})\|_n^\Theta < \infty$  by the same argument as before. For  $0 < n < 1$  the function  $\|\cdot\|_n$  is only a pseudo-norm as it is not sub-additive. However, the proof still follows by working instead with the metric  $\|\cdot\|_n^* := (\|\cdot\|_n)^n$  which is sub-additive; see the  $C_n$  inequality in Loève (1977).

The moments of the limit process  $x_t(\boldsymbol{\theta})$  are obtained as in the proof of Proposition TA.5. We note in particular that the limit of a sequence satisfying  $\|x_t(\cdot, \bar{x})\|_m^\Theta < \infty \forall t$ , that converges almost surely  $\sup_\theta \|x_t(\cdot, \bar{x}) - x_t(\cdot)\| \xrightarrow{a.s.} 0$ , and is uniformly integrable  $\lim_{M \rightarrow \infty} \sup_t \mathbb{E}[\|x_t(\cdot, \bar{x})\|_m^\Theta \cdot \mathbf{1}_{(|x_t(\cdot, \bar{x})| > M)}] = 0$  satisfies, by Theorem 2.17 in Jiang (2010),  $\mathbb{E}\|x_t(\cdot)\|_m^\Theta < \infty$ . In our case, boundedness of the moments of the initialized sequence has already been shown, and almost sure convergence is implied by exponentially fast almost sure convergence. We therefore only need to show the uniform integrability condition holds, which we can do for  $m = n - \delta$  for arbitrarily small  $\delta > 0$  by adapting Theorem 12.10 in Davidson (1994), since

$$\begin{aligned} \infty > \sup_t \left( \|x_t(\cdot, \bar{x})\|_n^\Theta \right)^n &= \sup_t \mathbb{E} \left[ \sup_{\boldsymbol{\theta} \in \Theta} |x_t(\boldsymbol{\theta}, \bar{x})|^n \right] \\ &\geq \sup_t \mathbb{E} \left[ \sup_{\boldsymbol{\theta} \in \Theta} |x_t(\boldsymbol{\theta}, \bar{x})|^n \mathbf{1}_{(|x_t(\boldsymbol{\theta}, \bar{x})| > M)} \right] \\ &\geq \sup_t \mathbb{E} \left[ \sup_{\boldsymbol{\theta} \in \Theta} |x_t(\boldsymbol{\theta}, \bar{x})|^m \mathbf{1}_{(|x_t(\cdot, \bar{x})| > M)} \sup_{\boldsymbol{\theta} \in \Theta} |x_t(\boldsymbol{\theta}, \bar{x})|^\delta \mathbf{1}_{(|x_t(\cdot, \bar{x})| > M)} \right] \\ &\geq M^\delta \sup_t \mathbb{E} \left[ \sup_{\boldsymbol{\theta} \in \Theta} |x_t(\boldsymbol{\theta}, \bar{x})|^m \mathbf{1}_{(|x_t(\cdot, \bar{x})| > M)} \right] \end{aligned}$$

Given the boundedness of the left-hand side of the equation and letting  $M \rightarrow \infty$ , we obtain that  $\lim_{M \rightarrow \infty} \sup_t \mathbb{E} \left[ \sup_{\boldsymbol{\theta} \in \Theta} |x_t(\boldsymbol{\theta}, \bar{x})|^m \mathbf{1}_{(|x_t(\cdot, \bar{x})| > M)} \right]$ , which establishes the result. ■

The following set of lemmas derives the bounds on the moments of the likelihood function based on moments of the inputs. The results follow from the properties of moment preserving maps as laid out in Technical Appendix F, but can also be proved directly.

**Lemma TA.11.**  $\mathbb{E} \sup_{\boldsymbol{\theta} \in \Theta} |\ell_T(\boldsymbol{\theta}, f)|^m < \infty$  where  $m \leq n_\nabla$  and for  $m \leq \min \{n_{\bar{g}}, n_{\bar{p}}\}$ .

*Proof.* The statement follows immediately from the fact that

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

■

**Lemma TA.12.**  $\mathbb{E} \sup_{\boldsymbol{\theta} \in \Theta} |\ell'_T(\boldsymbol{\theta}, f_1)|^m < \infty$  where

$$m = \min \left\{ n_{\bar{p}}^\lambda, \frac{n_\nabla n_{f_\theta}}{n_\nabla + n_{f_\theta}} \right\}. \quad (\text{E.5})$$

*Proof.* Using the explicit form of the first derivative of the likelihood in (D.2) in Technical Appendix D, the number of moments for the likelihood score is at least the minimum of the number of moments for each of the terms making up the score, namely

$$\frac{\partial \bar{p}}{\partial \boldsymbol{\theta}}, \quad \frac{\partial f_t}{\partial \boldsymbol{\theta}} \nabla_t.$$

The number of moments for the first term is  $n_{\bar{p}}^\lambda$ . Using a generalized Hölder inequality, the second term has moments  $n_\nabla n_{f_\theta} / (n_\nabla + n_{f_\theta})$ . This yields the expression for  $m$  in equation (E.5). ■



**Lemma TA.13.**  $\mathbb{E} \sup_{\boldsymbol{\theta} \in \Theta} |\ell_T''(\boldsymbol{\theta}, f)|^m < \infty$  where

$$m = \min \left\{ n_{\bar{p}}^{\lambda\lambda}, \frac{n_{\nabla} n_{f\boldsymbol{\theta}\boldsymbol{\theta}}}{n_{\nabla} + n_{f\boldsymbol{\theta}\boldsymbol{\theta}}}, \frac{n_{\nabla}^{\lambda} n_{f\boldsymbol{\theta}}}{n_{\nabla}^{\lambda} + n_{f\boldsymbol{\theta}}}, \frac{n_{\nabla}^f n_{f\boldsymbol{\theta}}}{2n_{\nabla}^f + n_{f\boldsymbol{\theta}}} \right\}. \quad (\text{E.6})$$

*Proof.* The statement follows by Holder's generalized inequality and from the explicit expression for the second derivative of the likelihood in equation (D.3) in Technical Appendix D, we obtain that the number of moments  $m$  is at least that of the minimum number of moments of the following terms

$$\frac{\partial^2 f_t}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^{\top}} \nabla_t, \quad \frac{\partial f_t}{\partial \boldsymbol{\theta}} \frac{\partial f_t}{\partial \boldsymbol{\theta}^{\top}} \frac{\partial \nabla_t}{\partial f_t}, \quad \frac{\partial f_t}{\partial \boldsymbol{\theta}} \frac{\partial \nabla_t}{\partial \lambda}, \quad \frac{\partial^2 \bar{p}_t}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^{\top}}.$$

Using generalized Hölder inequalities, the number of moments for each of these terms are, respectively,

$$\frac{n_{\nabla} n_{f\boldsymbol{\theta}\boldsymbol{\theta}}}{n_{\nabla} + n_{f\boldsymbol{\theta}\boldsymbol{\theta}}}, \quad \frac{n_{\nabla}^f n_{f\boldsymbol{\theta}}}{2n_{\nabla}^f + n_{f\boldsymbol{\theta}}}, \quad \frac{n_{\nabla}^{\lambda} n_{f\boldsymbol{\theta}}}{n_{\nabla}^{\lambda} + n_{f\boldsymbol{\theta}}}, \quad n_{\bar{p}}^{\lambda\lambda}.$$

This makes up the expression for  $m$  in equation (E.6). ■

The following lemmas support the proof of Theorem 4.14.

**Lemma TA.14.** *Let the conditions of Theorem 4.14 hold. Then  $\ell_T^l(\boldsymbol{\theta}_0)$  is a sample average of a sequence that is SE and NED of size  $-1$  on a strongly mixing sequence of size  $-\delta/(1-\delta)$  for some  $\delta > 2$ .*

*Proof.* By assumption,  $\{y_t\}_{t \in \mathbb{Z}}$  satisfies  $\mathbb{E}|y_t|^{n_y} < \infty$  for some  $n_y \geq 0$  and is SE and NED of size  $-1$  on a strongly mixing process of size  $-\delta/(1-\delta)$  for some  $\delta > 2$ . The uniform filter contraction in Assumption 4.4(a) and the moment conditions of Assumption 4.5 ensure that the limit process  $\{\mathbf{f}_t^{(0:1)}(\boldsymbol{\theta}_0)\}_{t \in \mathbb{Z}}$  is both SE (Proposition 3.5) and NED (Pötscher and Prucha (1997, Theorem 6.10)) of size  $-1$  on the strongly mixing process. The SE nature of the terms  $\tilde{\ell}'_t(y_t, \mathbf{f}_t^{(0:1)}(\boldsymbol{\theta}); \lambda)$  that compose the score

$$\ell_T'(\boldsymbol{\theta}, \bar{\mathbf{f}}_1) = \frac{1}{T} \sum_{t=1}^T \tilde{\ell}'_t(y_t, \mathbf{f}_t^{(0:1)}(\boldsymbol{\theta}); \lambda) = \frac{1}{T} \sum_{t=1}^T \frac{\partial f_t}{\partial \boldsymbol{\theta}} \cdot A_t^* + \frac{\partial \bar{p}_t}{\partial \boldsymbol{\theta}}$$

follows immediately by Krengel's theorem (Krengel (1985)) and the continuity of the score on the SE processes  $\{y_t\}_{t \in \mathbb{Z}}$  and  $\{\mathbf{f}_t^{(0:1)}(\boldsymbol{\theta}_0)\}_{t \in \mathbb{Z}}$ . Finally, the NED nature of the terms in  $\ell_T'(\boldsymbol{\theta}_0)$  is ensured by noting that Assumption 4.13 ensures that  $A_t^*$  is uniformly bounded, and hence that  $\tilde{\ell}'_t$  is Lipschitz continuous on  $(y_t, \mathbf{f}_t^{(0:1)}(\boldsymbol{\theta}_0))$ , which implies by Theorem 17.12 of Davidson (1994) or Theorem 6.15 of Pötscher and Prucha (1997) that  $\{\tilde{\ell}'_t(y_t, \mathbf{f}_t^{(0:1)}(\boldsymbol{\theta}); \lambda)\}$  is NED of size  $-1$  on the mixing sequence. ■

**Lemma TA.15.** *Under the conditions of Theorem 4.14,*

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

*Proof.* We establish the a.s. convergence in (E.7) by showing the e.a.s. convergence

$$\|\ell'_T(\boldsymbol{\theta}_0, \mathbf{f}_1^{(0:1)}) - \ell'_T(\boldsymbol{\theta}_0)\| \xrightarrow{e.a.s.} 0 \quad \text{as } T \rightarrow \infty.$$

This e.a.s. convergence follows from  $|f_t(\boldsymbol{\theta}_0, \bar{f}_1) - f_t(\boldsymbol{\theta}_0)| \xrightarrow{e.a.s.} 0$  and

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

as implied by Proposition 3.5 under the maintained assumptions. From the differentiability of

$$\tilde{\ell}'_t(\boldsymbol{\theta}, \mathbf{f}_1^{(0:1)}) = \ell'(\boldsymbol{\theta}, \mathbf{f}_t^{(0:1)}(\boldsymbol{\theta}, \mathbf{f}_1^{(0:1)}))$$

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

$$\begin{aligned} \|\ell'_T(\boldsymbol{\theta}_0, \mathbf{f}_1^{(0:1)}) - \ell'_T(\boldsymbol{\theta}_0)\| &\leq \sum_{j=1}^{4+d_\lambda} \left| \frac{\partial \ell'_t(\hat{\mathbf{f}}_t^{(0:1)})}{\partial f_j} \right| |\mathbf{f}_{j,t}^{(0:1)}(\boldsymbol{\theta}_0, \mathbf{f}_1^{(0:1)}) - \mathbf{f}_{j,t}^{(0:1)}(\boldsymbol{\theta}_0)| \\ &\leq \sum_{j=1}^{4+d_\lambda} \sup_{\mathbf{f}^{(0:1)}} \left| \frac{\partial \ell'_t(\mathbf{f}_t^{(0:1)})}{\partial f_j} \right| |\mathbf{f}_{j,t}^{(0:1)}(\boldsymbol{\theta}_0, \mathbf{f}_1^{(0:1)}) - \mathbf{f}_{j,t}^{(0:1)}(\boldsymbol{\theta}_0)|, \end{aligned} \quad (\text{E.8})$$

where  $\mathbf{f}_{j,t}^{(0:1)}$  denotes the  $j$ -th element of  $\mathbf{f}_t^{(0:1)}$ , and  $\hat{\mathbf{f}}^{(0:1)}$  is on the segment connecting  $\mathbf{f}_{j,t}^{(0:1)}(\boldsymbol{\theta}_0, \mathbf{f}_1^{(0:1)})$  and  $\mathbf{f}_{j,t}^{(0:1)}$ . Note that  $\mathbf{f}_t^{(0:1)} \in \mathbb{R}^{4+d_\lambda}$  because it contains  $f_t \in \mathbb{R}$  as well as  $\mathbf{f}_t^{(1)} \in \mathbb{R}^{3+d_\lambda}$ . Note that  $\sup_{\mathbf{f}^{(0:1)}} |\partial \ell'(\hat{\mathbf{f}}_t^{(0:1)}) / \partial f|$  has a logarithmic moment since  $\bar{n}_{\nabla}^f > 0$ . The strong convergence in (E.8) is now ensured by Lemma 2.1 in SM06 since

$$\|\ell'_T(\boldsymbol{\theta}_0, \mathbf{f}_1^{(0:1)}) - \ell'_T(\boldsymbol{\theta}_0)\| = \sum_{i=1}^{4+d_\lambda} \sup_{\mathbf{f}^{(0:1)}} \left| \frac{\partial \ell'_t(\mathbf{f}_t^{(0:1)})}{\partial f_j} \right| o_{e.a.s.}(1) = o_{e.a.s.}(1). \quad (\text{E.9})$$

■

**Lemma TA.16.** *Under the conditions of Theorem 4.14,  $\sup_{\boldsymbol{\theta} \in \Theta} \|\tilde{\ell}''_t(\boldsymbol{\theta}, \bar{f}_1) - \tilde{\ell}''_t(\boldsymbol{\theta})\| \xrightarrow{a.s.} 0$  as  $t \rightarrow \infty$ .*

*Proof.* The proof of the uniform convergence in (iv) is similar to that of Theorem 4.3. We note

$$\sup_{\boldsymbol{\theta} \in \Theta} \|\ell''_T(\boldsymbol{\theta}, \bar{f}_1) - \ell''_\infty(\boldsymbol{\theta})\| \leq \sup_{\boldsymbol{\theta} \in \Theta} \|\ell''_T(\boldsymbol{\theta}, \bar{f}_1) - \ell''_T(\boldsymbol{\theta})\| + \sup_{\boldsymbol{\theta} \in \Theta} \|\ell''_T(\boldsymbol{\theta}) - \ell''_\infty(\boldsymbol{\theta})\|. \quad (\text{E.10})$$

The first term vanishes e.a.s. by application of Lemma 2.1 in SM06 since  $\sup_{\boldsymbol{\theta} \in \Theta} \|\mathbf{f}_t^{(0:2)}(\boldsymbol{\theta}, \mathbf{f}_{0:2}) - \mathbf{f}_t^{(0:2)}(\boldsymbol{\theta})\| \xrightarrow{e.a.s.} 0$  and  $\sup_{\boldsymbol{\theta} \in \Theta} \sup_f \|\ell''_T(\boldsymbol{\theta})\|$  has a logarithmic moment since  $\bar{n}_{\nabla}^{ff} > 0$ .

The second term in (E.10) converges under a bound  $\mathbb{E} \sup_{\boldsymbol{\theta} \in \Theta} \|\tilde{\ell}''_t(\boldsymbol{\theta})\| < \infty$  by the SE nature of  $\{\ell''_T\}_{t \in \mathbb{Z}}$ . The latter is implied by continuity of  $\ell''$  on the SE sequence  $\{(y_t, \mathbf{f}_t^{(0:2)}(\cdot))\}_{t \in \mathbb{Z}}$  and Proposition 4.3 in Krengel (1985), where SE of  $\{(y_t, \mathbf{f}_t^{(0:2)}(\cdot))\}_{t \in \mathbb{Z}}$  follows from Proposition 3.5 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 4.11 and Lemma TA.5 in the Technical Appendix.

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

■

**Lemma TA.17.** Let  $\{x_t(\boldsymbol{\theta}, \bar{x})\}_{t \in \mathbb{N}}$  be a sequence initialized at  $\bar{x}$  that converges e.a.s. to an SE limit sequence  $\{x_t(\boldsymbol{\theta})\}_{t \in \mathbb{Z}}$ , i.e.,

$$|x_t(\boldsymbol{\theta}, \bar{x}) - x_t(\boldsymbol{\theta})| \xrightarrow{e.a.s.} 0 \quad \text{as } t \rightarrow \infty.$$

Let  $\mathbb{E} \log |x_t(\boldsymbol{\theta})| < \infty$ . Then

$$|x_t(\boldsymbol{\theta}, \bar{x})^2 - x_t(\boldsymbol{\theta})^2| \xrightarrow{e.a.s.} 0 \quad \text{as } t \rightarrow \infty.$$

*Proof.* By norm sub-additivity we have

$$\begin{aligned} |x_t(\boldsymbol{\theta}, \bar{x})^2 - x_t(\boldsymbol{\theta})^2| &= |x_t(\boldsymbol{\theta}, \bar{x}) - x_t(\boldsymbol{\theta})| \cdot |x_t(\boldsymbol{\theta}, \bar{x}) + x_t(\boldsymbol{\theta})| \\ &\leq |x_t(\boldsymbol{\theta}, \bar{x}) - x_t(\boldsymbol{\theta})| \cdot |x_t(\boldsymbol{\theta}, \bar{x}) - x_t(\boldsymbol{\theta}) + 2x_t(\boldsymbol{\theta})| \\ &\leq |x_t(\boldsymbol{\theta}, \bar{x}) - x_t(\boldsymbol{\theta})|^2 + 2|x_t(\boldsymbol{\theta}, \bar{x}) - x_t(\boldsymbol{\theta})| \cdot |x_t(\boldsymbol{\theta})|. \end{aligned}$$

The first term goes to zero e.a.s. due to the e.a.s. convergence of  $\{x_t(\boldsymbol{\theta}, \bar{x})\}_{t \in \mathbb{N}}$  to  $\{x_t(\boldsymbol{\theta})\}_{t \in \mathbb{Z}}$ . The second term goes to zero due to Lemma 2.1 in Straumann and Mikosch (2006), the e.a.s. convergence of  $\{x_t(\boldsymbol{\theta}, \bar{x})\}_{t \in \mathbb{N}}$  to  $\{x_t(\boldsymbol{\theta})\}_{t \in \mathbb{Z}}$ , and the SE nature and existence of a log moment of  $x_t(\boldsymbol{\theta})$ . ■

**Lemma TA.18.** Let  $\{\tilde{x}_t(\boldsymbol{\theta}, \bar{\tilde{x}})\}_{t \in \mathbb{N}}$  and  $\{x_t(\boldsymbol{\theta}, \bar{x})\}_{t \in \mathbb{N}}$  be sequences that converges e.a.s. to their SE limits  $\{\tilde{x}_t(\boldsymbol{\theta})\}_{t \in \mathbb{Z}}$  and  $\{x_t(\boldsymbol{\theta})\}_{t \in \mathbb{Z}}$ , respectively, i.e.,

$$|\tilde{x}_t(\boldsymbol{\theta}, \bar{\tilde{x}}) - \tilde{x}_t(\boldsymbol{\theta})| \xrightarrow{e.a.s.} 0, \quad |x_t(\boldsymbol{\theta}, \bar{x}) - x_t(\boldsymbol{\theta})| \xrightarrow{e.a.s.} 0 \quad \text{as } t \rightarrow \infty.$$

Let  $\mathbb{E} \log |\tilde{x}_t(\boldsymbol{\theta})| < \infty$  and  $\mathbb{E} \log |x_t(\boldsymbol{\theta})| < \infty$ . Then

$$|\tilde{x}_t(\boldsymbol{\theta}, \bar{\tilde{x}})x_t(\boldsymbol{\theta}, \bar{x}) - \tilde{x}_t(\boldsymbol{\theta})x_t(\boldsymbol{\theta})| \xrightarrow{e.a.s.} 0 \quad \text{as } t \rightarrow \infty.$$

*Proof.* We have

$$\begin{aligned} &|\tilde{x}_t(\boldsymbol{\theta}, \bar{\tilde{x}})x_t(\boldsymbol{\theta}, \bar{x}) - \tilde{x}_t(\boldsymbol{\theta})x_t(\boldsymbol{\theta})| \\ &= |\tilde{x}_t(\boldsymbol{\theta}, \bar{\tilde{x}})x_t(\boldsymbol{\theta}, \bar{x}) - \tilde{x}_t(\boldsymbol{\theta}, \bar{\tilde{x}})x_t(\boldsymbol{\theta}) + \tilde{x}_t(\boldsymbol{\theta}, \bar{\tilde{x}})x_t(\boldsymbol{\theta}) - \tilde{x}_t(\boldsymbol{\theta})x_t(\boldsymbol{\theta})| \\ &\leq |\tilde{x}_t(\boldsymbol{\theta}, \bar{\tilde{x}})| \cdot |x_t(\boldsymbol{\theta}, \bar{x}) - x_t(\boldsymbol{\theta})| + |\tilde{x}_t(\boldsymbol{\theta}, \bar{\tilde{x}}) - \tilde{x}_t(\boldsymbol{\theta})| \cdot |x_t(\boldsymbol{\theta})| \\ &\leq |\tilde{x}_t(\boldsymbol{\theta}, \bar{\tilde{x}}) - \tilde{x}_t(\boldsymbol{\theta}) + \tilde{x}_t(\boldsymbol{\theta})| \cdot |x_t(\boldsymbol{\theta}, \bar{x}) - x_t(\boldsymbol{\theta})| + \\ &\quad |\tilde{x}_t(\boldsymbol{\theta}, \bar{\tilde{x}}) - \tilde{x}_t(\boldsymbol{\theta})| \cdot |x_t(\boldsymbol{\theta})| \\ &\leq |\tilde{x}_t(\boldsymbol{\theta}, \bar{\tilde{x}}) - \tilde{x}_t(\boldsymbol{\theta})| \cdot |x_t(\boldsymbol{\theta}, \bar{x}) - x_t(\boldsymbol{\theta})| + \\ &\quad |\tilde{x}_t(\boldsymbol{\theta})| \cdot |x_t(\boldsymbol{\theta}, \bar{x}) - x_t(\boldsymbol{\theta})| + \\ &\quad |\tilde{x}_t(\boldsymbol{\theta}, \bar{\tilde{x}}) - \tilde{x}_t(\boldsymbol{\theta})| \cdot |x_t(\boldsymbol{\theta})| \end{aligned}$$

The first term goes to zero e.a.s. due to the e.a.s. convergence of  $\{\tilde{x}_t(\boldsymbol{\theta}, \bar{\tilde{x}})\}_{t \in \mathbb{N}}$  and  $\{x_t(\boldsymbol{\theta}, \bar{x})\}_{t \in \mathbb{N}}$  to  $\{\tilde{x}_t(\boldsymbol{\theta})\}_{t \in \mathbb{Z}}$  and  $\{x_t(\boldsymbol{\theta})\}_{t \in \mathbb{Z}}$ , respectively. The second and third term go to zero due to Lemma 2.1 in Straumann and Mikosch (2006), the e.a.s. convergence of  $\{\tilde{x}_t(\boldsymbol{\theta}, \bar{\tilde{x}})\}_{t \in \mathbb{N}}$  and  $\{x_t(\boldsymbol{\theta}, \bar{x})\}_{t \in \mathbb{N}}$ , and the SE nature and existence of a log moment for both  $\tilde{x}_t(\boldsymbol{\theta})$  and  $x_t(\boldsymbol{\theta})$ . ■

**Lemma TA.19.** Let  $\mathbf{A}_t^{(2)}(\boldsymbol{\theta}, \bar{\mathbf{f}}_1^{0:1})$  be as defined in (D.6) and evaluated at the initialized series for  $f_t(\boldsymbol{\theta}, \bar{f}_1)$  and  $\mathbf{f}_t^{(1)}(\boldsymbol{\theta}, \bar{\mathbf{f}}_1^{0:1})$ . Similarly, let  $\mathbf{A}_t^{(2)}(\boldsymbol{\theta})$  denote the same quantity evaluated at the SE limits  $f_t(\boldsymbol{\theta})$  and  $\mathbf{f}_t^{(1)}(\boldsymbol{\theta})$ . Then under the conditions of Proposition 3.5, we have

$$|\mathbf{A}_t^{(2)}(\boldsymbol{\theta}, \bar{\mathbf{f}}_1^{0:1}) - \mathbf{A}_t^{(2)}(\boldsymbol{\theta})| \xrightarrow{e.a.s.} 0.$$

*Proof.* Under the conditions of Proposition 3.5  $f_t(\boldsymbol{\theta}, \bar{f}_1) \xrightarrow{e.a.s.} f_t(\boldsymbol{\theta})$  and  $\mathbf{f}_t^{(1)}(\boldsymbol{\theta}, \bar{\mathbf{f}}_1^{0:1}) \xrightarrow{e.a.s.} \mathbf{f}_t^{(1)}(\boldsymbol{\theta})$ . The expression for  $\mathbf{A}_t^{(2)}$  in (D.6) has three different types of terms.

*Type I:* The terms

$$\frac{\partial \alpha}{\partial \boldsymbol{\theta}} \frac{\partial s_t}{\partial \boldsymbol{\theta}^\top}, \quad \alpha \frac{\partial^2 s_t}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top}, \quad \frac{\partial \beta}{\partial \boldsymbol{\theta}} \frac{\partial f_t}{\partial \boldsymbol{\theta}^\top}.$$

These terms for  $|\mathbf{A}_t^{(2)}(\boldsymbol{\theta}, \bar{\mathbf{f}}_1^{0:1}) - \mathbf{A}_t^{(2)}(\boldsymbol{\theta})|$  converge e.a.s. to zero.

The first term follows by noting that  $\partial \alpha / \partial \boldsymbol{\theta}$  is constant, and

$$\left| \frac{\partial s_t(f_t(\boldsymbol{\theta}, \bar{f}_1))}{\partial \boldsymbol{\theta}} - \frac{\partial s_t(f_t(\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}} \right| \leq \left( \sup_{f^*} \left| \frac{\partial^2 s_t(f^*)}{\partial f \partial \boldsymbol{\theta}} \right| \right) |f_t(\boldsymbol{\theta}, \bar{f}_1) - f_t(\boldsymbol{\theta})|. \quad (\text{E.11})$$

The result now follows from Lemma 2.1 in Straumann and Mikosch (2006) due to the e.a.s. convergence  $f_t(\boldsymbol{\theta}, \bar{f}_1) \xrightarrow{e.a.s.} f_t(\boldsymbol{\theta})$ , the SE nature of the term involving the sup, and the existence of a small positive moment for the sup, which implies the existence of a log moment.

The e.a.s. convergence for the second term follows by a similar argument.

The third term follows directly from the e.a.s. convergence  $\mathbf{f}_t^{(1)}(\boldsymbol{\theta}, \bar{\mathbf{f}}_1^{0:1}) \xrightarrow{e.a.s.} \mathbf{f}_t^{(1)}(\boldsymbol{\theta})$ .

*Type II:* The terms

$$\frac{\partial \alpha}{\partial \boldsymbol{\theta}} \frac{\partial s_t}{\partial f_t} \frac{\partial f_t}{\partial \boldsymbol{\theta}^\top}, \quad \alpha \frac{\partial^2 s_t}{\partial \boldsymbol{\theta} \partial f_t} \frac{\partial f_t}{\partial \boldsymbol{\theta}^\top}.$$

Both terms follow by a similar argument as the first set of terms, combined with Lemma TA.18. For instance for the first term, we have  $\partial \alpha / \partial \boldsymbol{\theta}$  is constant, and

$$\frac{\partial s_t(f_t(\boldsymbol{\theta}, \bar{f}_1))}{\partial \boldsymbol{\theta}} \xrightarrow{e.a.s.} \frac{\partial s_t(f_t(\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}} \quad (\text{E.12})$$

given the arguments under terms of Type I. Given the e.a.s. convergence of both  $\partial s_t(f_t(\boldsymbol{\theta}, \bar{f}_1)) / \partial \boldsymbol{\theta}$  and  $f_t(\boldsymbol{\theta}, \bar{f}_1)$ , the results follow directly from Lemma TA.18.

*Type III:* The term

$$\alpha \frac{\partial^2 s_t}{\partial f_t^2} \frac{\partial f_t}{\partial \boldsymbol{\theta}} \frac{\partial f_t}{\partial \boldsymbol{\theta}^\top}.$$

The existence of a log moment for  $\sup_{f^*} |\partial^3 s_t / \partial f_t^3|$  is implied by  $\bar{n}_s^{fff} > 0$ . This again implies the e.a.s. convergence of  $\partial^2 s_t / \partial f_t^2$  via Lemma 2.1 in Straumann and Mikosch (2006). The e.a.s. convergence of each of the elements in  $(\partial f_t / \boldsymbol{\theta})(\partial f_t / \boldsymbol{\theta}^\top)$  follows from Lemmas TA.17 and TA.18 given a log moment for each of the elements of  $\partial f_t / \boldsymbol{\theta}$ , which is implied by  $n_{f_\theta} > 0$ . Note that the latter also implies a log moment for  $(\partial f_t / \boldsymbol{\theta})(\partial f_t / \boldsymbol{\theta}^\top)$ . We also have a log moment for  $f_t(\boldsymbol{\theta})$  given  $n_f > 0$ . Combining all these results, we have

$$\begin{aligned} & \left| \frac{\partial^2 s_t(f_t(\boldsymbol{\theta}, \bar{f}_1))}{\partial f_t^2} \frac{\partial f_t(\boldsymbol{\theta}, \bar{\mathbf{f}}_1^{0:1})}{\partial \boldsymbol{\theta}_i} \frac{\partial f_t(\boldsymbol{\theta}, \bar{\mathbf{f}}_1^{0:1})}{\partial \boldsymbol{\theta}_j} - \frac{\partial^2 s_t(f_t(\boldsymbol{\theta}))}{\partial f_t^2} \frac{\partial f_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i} \frac{\partial f_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j} \right| \\ & \leq \left| \sup_{f^*} \frac{\partial^3 s_t(f^*)}{\partial f_t^3} \right| \times |f_t(\boldsymbol{\theta}, \bar{f}_1) - f_t(\boldsymbol{\theta})| \times \\ & \quad \left| \frac{\partial f_t(\boldsymbol{\theta}, \bar{\mathbf{f}}_1^{0:1})}{\partial \boldsymbol{\theta}_i} \frac{\partial f_t(\boldsymbol{\theta}, \bar{\mathbf{f}}_1^{0:1})}{\partial \boldsymbol{\theta}_j} - \frac{\partial f_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i} \frac{\partial f_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j} \right|. \end{aligned}$$

This expression contains three factors. We already concluded that the second and third factor converge e.a.s. to zero and have a log moment, such that also their product converges e.a.s. to zero using Lemma TA.18, and also has a log moment. The latter follows because under the current assumptions each of the two elements of the product actually has a (tiny) positive moment, which implies a log moment for the product. Multiplying the last two factors by the first factor, and noting that the first factor is SE and has a log moment under the current set of assumptions, the result follows directly from Lemma 2.1 in Straumann and Mikosch (2006). ■

## F Further Results on Moment Preserving Functions

Checking the moment conditions needed for a number of the propositions and theorems based on low-level conditions can be considerably simplified by working with the concept of moment preserving maps.

The final technical lemma presented below provides simple moment preserving properties for several common functions of random variables. For notational simplicity we let  $h^{(k)}$  denote the  $k$ th order derivative of a function  $h$ . The moment properties on  $h$  or  $h^{(k)}$  can now easily be derived from moment conditions on the inputs of  $h$  and the moment preserving properties through its membership of the set  $\mathbb{M}_{\Theta, \Theta}(n, m)$ .

**Lemma TA.20.** (Catalog of  $\mathbb{M}_{\Theta, \Theta}^k(n, m)$  Moment Preserving Maps) *For every  $\theta \in \Theta$ , let  $h(\cdot; \theta) : \mathcal{X} \rightarrow \mathbb{R}$  and  $w(\cdot, \cdot, \theta) : \mathcal{X} \times \mathcal{V} \rightarrow \mathbb{R}$  be measurable functions.*

(a) *Let  $h(\cdot; \theta)$  be an affine function,*

$$h(x; \theta) = \theta_0 + \theta_1 x \quad \forall (x, \theta) \in \mathcal{X} \times \Theta, \quad \theta = (\theta_0, \theta_1) \in \Theta \subseteq \mathbb{R}^2.$$

*Then,  $h(\cdot; \theta) \in \mathbb{M}_{\Theta, \theta}(n, m)$  with  $n = m \quad \forall \theta \in \Theta$ , and  $h^{(k)}(\cdot; \theta) \in \mathbb{M}_{\Theta, \theta}(n, m)$  for all  $(\theta, n, m, k) \in \Theta \times \mathbb{R}_0^+ \times \mathbb{R}_0^+ \times \mathbb{N}$ . If  $\Theta$  is compact, then  $h \in \mathbb{M}_{\Theta, \Theta}^k(n, m)$  with  $n = m$  for  $k = 0$  and  $h^{(k)}(\cdot; \theta) \in \mathbb{M}_{\Theta, \Theta}(n, m) \quad \forall (n, m, k) \in \mathbb{R}_0^+ \times \mathbb{R}_0^+ \times \mathbb{N}$ .*

(b) *Let  $h(\cdot; \theta)$  be a polynomial function,*

$$h(x; \theta) = \sum_{j=0}^J \theta_j x^j \quad \forall (x, \theta) \in \mathcal{X} \times \Theta, \quad \theta = (\theta_0, \dots, \theta_J) \in \Theta \subseteq \mathbb{R}^J, J \geq 1.$$

*Then  $h^{(k)}(\cdot; \theta) \in \mathbb{M}_{\Theta, \theta}(n, m)$  with  $m = n/(J - k) \quad \forall (k, \theta) \in \mathbb{N}_0 \times \Theta$ . If  $\Theta$  is compact, then  $h^{(k)} \in \mathbb{M}_{\Theta, \Theta}(n, m)$  with  $m = n/(J - k) \quad \forall k \in \mathbb{N}_0$ .*

(c) *Let*

$$h(x; \theta) = \sum_{j=0}^J \theta_j x^{r_j} \quad \forall (x, \theta) \in \mathcal{X} \times \Theta, \quad \theta = (\theta_0, \dots, \theta_J) \in \Theta \subseteq \mathbb{R}^J,$$

*where  $r_j \geq 0$ . Then  $h^{(k)}(\cdot; \theta) \in \mathbb{M}_{\Theta, \theta}(n, m)$  with  $m = n/(\max_j r_j - k) \quad \forall (\theta, k) \in \Theta \in \mathbb{N}_0 : k \leq \min_j r_j$ . If  $\Theta$  is compact, then  $h^{(k)} \in \mathbb{M}_{\Theta, \Theta}(n, m)$  with  $m = n/(\max_j r_j - k) \quad \forall k \in \mathbb{N}_0 : k \leq \min_j r_j$ .*

(d) *Let*

$$\sup_{x \in \mathcal{X}} |h(x; \theta)| \leq \bar{h}(\theta) < \infty \quad \forall \theta \in \Theta.$$

*Then  $h(\cdot; \theta) \in \mathbb{M}_{\Theta, \theta}(n, m) \quad \forall (n, m, \theta) \in \Theta \times \mathbb{R}_0^+ \times \mathbb{R}_0^+$ . If additionally,  $\sup_{\theta \in \Theta} \bar{h}(\theta) \leq \bar{\bar{h}} < \infty$ , then  $h \in \mathbb{M}_{\Theta, \Theta}(n, m) \quad \forall (n, m) \in \mathbb{R}_0^+ \times \mathbb{R}_0^+$ .*

(e) Let

$$h(\cdot; \boldsymbol{\theta}) \in \mathbb{C}^k(\mathcal{X})$$

and

$$\sup_{x \in \mathcal{X}} |h^{(k)}(x; \boldsymbol{\theta})| \leq \bar{h}_k(\boldsymbol{\theta}) < \infty \quad \forall \boldsymbol{\theta} \in \Theta.$$

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

(f) Let

$$w(x, v; \boldsymbol{\theta}) = \theta_0 + \theta_1 x + \theta_2 v, \quad (\theta_0, \theta_1, \theta_2, x, v) \in \mathbb{R}^3 \times \mathcal{X} \times \mathcal{V}.$$

Then  $w^{(k_x, k_v)}(\cdot, \cdot, \boldsymbol{\theta}) \in \mathbb{M}_{\Theta, \boldsymbol{\theta}}(\mathbf{n}, m) \quad \forall (k_x, k_v, \boldsymbol{\theta}) \in \mathbb{N}_0 \times \mathbb{N}_0 \times \Theta$  with  $\mathbf{n} = (n_x, n_v)$  and  $m = \min\{n_x, n_v\}$ . If furthermore  $\Theta$  is compact, then

$$w^{(k_x, k_v)} \in \mathbb{M}_{\Theta, \Theta}(\mathbf{n}, m) \quad \forall (k_x, k_v) \in \mathbb{N}_0 \times \mathbb{N}_0,$$

with  $m = \min\{n_x, n_v\}$ ;

(g) If

$$w(x, v, \boldsymbol{\theta}) = \theta_0 + \theta_1 x v, \quad (\theta_0, \theta_1) \in \mathbb{R}^2,$$

then  $w^{(k_x, k_v)}(\cdot, \cdot, \boldsymbol{\theta}) \in \mathbb{M}_{\Theta, \boldsymbol{\theta}}(\mathbf{n}, m) \quad \forall (k_x, k_v, \boldsymbol{\theta}) \in \mathbb{N}_0 \times \mathbb{N}_0 \times \Theta$  with  $\mathbf{n} = (n_x, n_v)$  where  $m = n_x n_v / (n_x + n_v)$ . If furthermore,  $\Theta$  is compact, then

$$w^{(k_x, k_v)} \in \mathbb{M}_{\Theta, \Theta}(\mathbf{n}, m) \quad \forall (k_x, k_v) \in \mathbb{N}_0 \times \mathbb{N}_0,$$

with  $\mathbf{n} = (n_x, n_v)$  where  $m = n_x n_v / (n_x + n_v)$ .

Proof available upon request.