

TI 2025-063/III
Tinbergen Institute Discussion Paper

Testing for the Absence of Score-Driven Parameter Dynamics

*Andre Lucas*¹

*Yicong Lin*²

¹ Vrije Universiteit Amsterdam, Tinbergen Institute

² Vrije Universiteit Amsterdam, Tinbergen Institute

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

Contact: discussionpapers@tinbergen.nl

More TI discussion papers can be downloaded at <https://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

TESTING FOR THE ABSENCE OF SCORE-DRIVEN PARAMETER DYNAMICS

Yicong Lin^{*1} and Andre Lucas¹

¹ Vrije Universiteit Amsterdam and Tinbergen Institute

October, 2025

Abstract

This paper proposes a quasi-likelihood ratio (QLR) test for the null of constant parameters against the alternative of score-driven parameter dynamics. Score-driven models have been widely used in the literature to capture time variation in parameters across a diverse range of both continuous and discrete, univariate and multivariate time series models, with or without random regressors. A formal testing procedure, however, is lacking thus far. Our QLR test addresses two key challenges: (i) parameters may lie on the boundary of the parameter space, and (ii) nuisance parameters are not identified under the null. The test statistic's non-standard asymptotic distribution takes a simple form that only depends on the specified parameter space and is invariant to the specific formulation of the score-driven model and its degree of nonlinearity. Consequently, the asymptotic distribution applies to a wide range of score-driven models and can easily be simulated to conduct inference. We illustrate the new test using several models from the score-driven literature and show that the limiting distribution provides an adequate approximation for inference in finite samples.

Keywords: parameter constancy, score-driven models, quasi-likelihood ratio test, parameters on the boundary, nonidentification.

^{*}Corresponding author: Department of Econometrics and Data Science, Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV, Amsterdam, the Netherlands. E-mail address: yc.lin@vu.nl.

1 Introduction

Observation-driven models are widely adopted for capturing time variation in time series data. In these models, the dynamics are specified as functions of past observations. The time-varying parameters are therefore pre-determined, though unconditionally stochastic. This feature of observation-driven models allows for a standard application of maximum likelihood methods. Among the class of observation-driven models, a prominent and extensively studied subclass is the family of *score-driven models*, introduced independently by [Creal, Koopman, and Lucas \(2013\)](#) and [Harvey \(2013\)](#).¹ The theoretical underpinnings of score-driven models have recently been examined in, among others, [Blasques, Koopman, and Lucas. \(2015, 2018\)](#), [Blasques, van Brummelen, Koopman, and Lucas \(2022\)](#), and [Gorgi, Lauria, and Luati \(2024\)](#). A particularly interesting result is that score-driven models are optimal within the class of observation-driven models, in the sense that they achieve the minimum asymptotic variance, even when the dynamics are not observation-driven ([Beutner, Lin, and Lucas, 2023](#)).

However, to the best of our knowledge, previous studies, whether theoretical or applied, generally assume the presence of score-driven time variation in the parameters from the outset. This assumption conveniently ensures unique parameter identification and enables standard asymptotic approximations for M -type estimators. If the assumption is violated, however, the corresponding asymptotic results typically fail, undermining the reliability of inference in practice. Despite the widespread use of score-driven models, the literature is surprisingly thin on formal testing procedures to detect time variation. Only [Calvori, Creal, Koopman, and Lucas \(2017\)](#) develop an LM testing procedure for score-driven parameter dynamics, but without any formal theory and relying on a rather restrictive solution to the nonidentification problem that arises under the null, as discussed later. Instead, most papers rely on ad-hoc or intuitive visual diagnostics like the size of the time variation in the filtered parameters. Such approaches are statistically unreliable and can even be misleading. In particular, when time variation is absent, naively applying score-driven models can produce strong, but spurious time-varying patterns. For this reason, a formal test for constant parameters against the alternative of score-driven dynamics is necessary, and this is the focus of the present paper.

We consider a flexible setup that accommodates different types of variables of interest, including, for example, discrete or continuous variables as well as random exogenous variables. We propose a quasi-likelihood ratio (QLR) test for the constancy of a scalar parameter against score-driven dynamics, allowing for potential model mis-specification. Two main challenges

¹Interested readers are referred to <https://www.gasmodel.com/>, which provides a collection of over 400 recent articles related to this line of research.

arise, however, when deriving the limiting distribution of the QLR test under the null. First, parameters may lie on the boundary of the parameter space under the null hypothesis, which invalidates conventional asymptotic analysis that typically requires parameters to lie in the interior of the parameter space (see, e.g., the pioneering works on (G)ARCH models [Francq and Zakoïan, 2009](#); [Cavaliere, Nielsen, and Rahbek, 2017](#); [Jiang, Li, and Zhu, 2020](#); [Cavaliere, Nielsen, Pedersen, and Rahbek, 2022](#); [Cavaliere, Perera, and Rahbek, 2024](#)). Second, the test runs into the well-known Davies problem ([Davies, 1977, 1987](#)), whereby some parameters are not identified under the null (see, e.g., [Andrews, 1993](#); [Andrews and Ploberger, 1994, 1996](#); [Hansen, 1996](#); [Baek, Cho, and Phillips, 2015](#)). The studies referred to above typically focus on one of these two issues. By contrast, we address both challenges at the same time, building on the seminal work of [Andrews \(2001\)](#) and deriving the limiting null distribution of the proposed QLR test. We establish a set of easily verifiable conditions and illustrate their application to a range of different popular score-driven models.

Our findings show that the limiting distribution generally depends on several nuisance parameters. However, under mild conditions that are typically satisfied by score-driven models from the literature, the QLR statistic converges, either directly or after rescaling, to a simple yet non-standard limiting distribution that is free of nuisance parameters. This limiting distribution does not follow the conventional χ^2 law or a mixture thereof (see [Kopylev and Sinha, 2011](#), and references therein), but instead depends on the specification of the parameter space. The limiting distribution can be readily simulated for practical implementation and inference. Interestingly, the same limiting distribution as obtained in this paper has appeared in earlier [Andrews and Ploberger \(1996\)](#) for tests of serial correlation in ARMA(1,1) models, and in [Andrews \(2001\)](#) for tests of conditional heteroskedasticity. The current paper shows that it applies to a much broader context of time-varying parameter models that are used in practice.

The remainder of the paper is organized as follows. Section 2 introduces the setup and presents the QLR test. Section 3 derives the asymptotic null distribution and discusses the conditions that allow for a nuisance parameter free limiting distribution. This section also provides additional results offering sufficient conditions that can be used to verify some of the main assumptions. Section 4 illustrates the theoretical results using different models. Section 5 presents simulation evidence to evaluate the performance of our asymptotic approximations in finite samples. Section 6 presents a short empirical application of the time-varying spatial regression model of [Blasques, Koopman, Lucas, and Schaumburg \(2016\)](#). Section 7 concludes. The Online Appendix contains all the proofs, a comprehensive set of simulated critical values, and further details on the examples in Section 4.

2 Testing for score-driven parameter dynamics

2.1 Score-driven time-varying parameter models

For $t \in \mathbb{Z}^+$, let $\mathbf{y}_t \in \mathcal{Y} \subset \mathbb{R}^n$ denote the vector of variables of interest, and $\mathbf{x}_t \in \mathcal{X} \subset \mathbb{R}^k$ a vector of exogenous variables, where $n \geq 1$ and $k \geq 0$, with $k = 0$ when no exogenous variables are present. Define $\mathcal{F}_t := \sigma((\mathbf{y}_s, \mathbf{x}_s), s \leq t)$, i.e., the σ -field generated by $\{(\mathbf{y}_t, \mathbf{x}_t), t \in \mathbb{Z}\}$. We consider the setting where the statistician assumes a conditional probability density function (pdf) or probability mass function (pmf) $p(\cdot \mid \mathbf{x}_t, f_t, \boldsymbol{\phi})$ for $\mathbf{y}_t$ given $\mathbf{x}_t$ and $\mathcal{F}_{t-1}$, where $f_t \in \mathcal{F} \subset \mathbb{R}$ is a scalar time-varying parameter that is of primary interest, and $\boldsymbol{\phi} \in \boldsymbol{\Theta}_\phi$ is a vector of static parameters. The assumed conditional pdf/pmf $p(\cdot \mid \mathbf{x}_t, f_t, \boldsymbol{\phi})$ does not necessarily represent the true, unknown conditional pdf/pmf. In other words, the model may be mis-specified. In this setting, [Creal et al. \(2013\)](#) and [Harvey \(2013\)](#) propose the popular class of score-driven parameter dynamics for f_t as given by

$$f_{t+1} = \psi_t(f_t, \boldsymbol{\vartheta}, \beta) = \omega(1 - \beta) + \beta f_t + \alpha s_t(f_t, \boldsymbol{\phi}), \quad (2.1)$$

$$s_t(f, \boldsymbol{\phi}) := s(\mathbf{y}_t, \mathbf{x}_t, f, \boldsymbol{\phi}) = S(\mathbf{y}_t, \mathbf{x}_t, f, \boldsymbol{\phi}) \frac{\partial \log p(\mathbf{y}_t \mid \mathbf{x}_t, f, \boldsymbol{\phi})}{\partial f}, \quad (2.2)$$

where $\boldsymbol{\vartheta} := (\alpha, \omega, \boldsymbol{\phi}^\top)^\top \in \boldsymbol{\Theta} = \Theta_\alpha \times \Theta_\omega \times \boldsymbol{\Theta}_\phi \subset \mathbb{R}^{m+2}$, $\beta \in \mathcal{B} \subset \mathbb{R}$, and $S(\mathbf{y}_t, \mathbf{x}_t, f, \boldsymbol{\phi}) \in \mathbb{R}$ is a user-specified scale. Note that we treat the parameter β and its parameter space $\mathcal{B}$ differently from the remaining parameters. The reason for this becomes apparent in the next subsection. Additional conditions on the parameter space $\boldsymbol{\Theta}$ are provided in Section 3. For convenience, we define the shorthand notation $S_t(f, \boldsymbol{\phi}) := S(\mathbf{y}_t, \mathbf{x}_t, f, \boldsymbol{\phi})$ and $\nabla_t^f(f, \boldsymbol{\phi}) := \partial \log p(\mathbf{y}_t \mid \mathbf{x}_t, f, \boldsymbol{\phi}) / \partial f$.

The score-driven time-varying parameter setup in (2.1)–(2.2) encompasses a wide range of empirically relevant models, including, among others, volatility models ([Creal, Koopman, and Lucas., 2011](#); [Harvey, 2013](#)), dynamic copula and correlation models ([Janus, Koopman, and Lucas, 2014](#); [Blasques, Lucas, and Silde, 2018](#); [Opschoor, Janus, Lucas, and Van Dijk, 2018](#)), integer time series models ([Fokianos, Rahbek, and Tjøstheim, 2009](#); [Babii, Chen, and Ghysels, 2019](#); [Gorgi, 2020](#)), robust location models ([Harvey and Luati, 2014](#)), dynamic spatial Durbin models ([Blasques et al., 2016](#); [D’Innocenzo, Lucas, Opschoor, and Zhang, 2024](#)), time-varying tail shape models ([Massacci, 2017](#); [D’Innocenzo, Lucas, Schwaab, and Zhang, 2024](#)), matrix-valued and tensor-valued time series ([Armiliotta, Gorgi, and Lucas, 2025](#); [Lin, Lucas, and Ye, 2025](#)), and more.

Even though the assumed pdf/pmf may be mis-specified, the score-driven parameter dynamics in (2.1)–(2.2) result in expected local improvements of the Kullback-Leibler divergence (Blasques et al., 2015, 2018; De Punder, Dimitriadis, and Lange, 2024; Gorgi et al., 2024). In addition, Beutner et al. (2023) show that the score-driven filtered values f_t provide consistent estimates of the true, unobserved Kullback-Leibler optimal parameter dynamics under various forms of severe mis-specification. Both results underline the theoretical underpinnings and flexibility of the score-driven approach.

2.2 Quasi-likelihood ratio (QLR) tests for parameter constancy

Despite the widespread use of the score-driven modeling methodology, there is to the best of our knowledge currently no formal test available for the presence of score-driven time-varying parameters. This is precisely the main focus of the current paper. Given the general asymptotic results in Beutner et al. (2023) on the Kullback-Leibler consistency of f_t under general forms of mis-specification, such a test may even be viewed as a general test of time variation in the true data generating process (dgp), whether the assumed pdf/pmf $p(\cdot \mid \mathbf{x}_t, f_t, \phi)$ is correctly specified or not.

Let $\{0\} \subset \Theta_\alpha$, then testing for time invariance reduces to testing the null hypothesis $H_0 : \alpha = 0$ against the alternative $H_1 : \alpha \neq 0$. Two remarks are in place. First, depending on the model, $\alpha = 0$ may lie on the boundary of Θ_α . For example, in several score-driven volatility models α is assumed to be non-negative in order to ensure positive variances for all t . If $\alpha = 0$ lies on the boundary of the parameter space, this requires a careful asymptotic treatment of the testing procedure (Andrews, 1999, 2001). Second, if $|\beta| < 1$, which is a standard assumption in the literature, the parameter β is absent under H_0 and thus becomes unidentified, similar to the situation considered in Andrews and Ploberger (1994, 1996). This invalidates standard testing procedures that do not have such identification issues. The issue is also known as the Davies problem, where a nuisance parameter is identifiable only under the alternative (Davies, 1977, 1987).

To define our test statistic, let $\ell_t(f, \phi) := \ell(\mathbf{y}_t, \mathbf{x}_t, f, \phi) = \log p(\mathbf{y}_t \mid \mathbf{x}_t, f, \phi)$ for $p(\mathbf{y}_t \mid \mathbf{x}_t, f, \phi) > 0$ and $-\infty$ otherwise. Moreover, let $\hat{f}_t(\boldsymbol{\vartheta}, \beta) := \hat{f}_t(\boldsymbol{\vartheta}, \beta; \hat{f}_1(\boldsymbol{\vartheta}, \beta))$ be an initialized filtering sequence. We define the empirical quasi-log likelihood $\hat{\mathcal{L}}_T$ and the unrestricted estimator $\hat{\boldsymbol{\vartheta}}_{\beta, T}$ as, $\forall \beta \in \mathcal{B}$,

$$\hat{\boldsymbol{\vartheta}}_{\beta, T} = \operatorname{argmax}_{\boldsymbol{\vartheta} \in \Theta} \hat{\mathcal{L}}_T(\boldsymbol{\vartheta}, \beta) + o_{\mathbb{P}}(1), \quad \hat{\mathcal{L}}_T(\boldsymbol{\vartheta}, \beta) := \sum_{t=1}^T \ell_t(\hat{f}_t(\boldsymbol{\vartheta}, \beta), \phi), \quad (2.3)$$

where the $o_{\mathbb{P}}(1)$ term is uniform in $\beta \in \mathcal{B}$. The estimator $\widehat{\boldsymbol{\vartheta}}_{\beta,T} = (\widehat{\alpha}_{\beta,T}, \widehat{\omega}_{\beta,T}, \widehat{\boldsymbol{\phi}}_{\beta,T}^{\top})^{\top}$ thus denotes the unrestricted extremum estimator for a given $\beta \in \mathcal{B}$, allowing the supremum to be attained approximately rather than exactly, with an approximation error of asymptotic order $o_{\mathbb{P}}(1)$. Similarly, the restricted estimator $\widehat{\boldsymbol{\vartheta}}_{0\beta,T}$ is defined by replacing the maximum over $\boldsymbol{\Theta}$ with that over $\boldsymbol{\Theta}_0$ in Eq. (2.3), where $\boldsymbol{\Theta}_0 = \{\boldsymbol{\vartheta} \in \boldsymbol{\Theta} : \boldsymbol{\vartheta} = (0, \omega, \boldsymbol{\phi}^{\top})^{\top}\}$ is the restricted parameter space under the null hypothesis. The QLR test statistic is then defined as

$$QLR_T = -2 \left(\sup_{\beta \in \mathcal{B}} \widehat{\mathcal{L}}_T(\widehat{\boldsymbol{\vartheta}}_{0\beta,T}, \beta) - \sup_{\beta \in \mathcal{B}} \widehat{\mathcal{L}}_T(\widehat{\boldsymbol{\vartheta}}_{\beta,T}, \beta) \right) + o_{\mathbb{P}}(1). \quad (2.4)$$

Let $\boldsymbol{\vartheta}_0 = (\alpha_0, \omega_0, \boldsymbol{\phi}_0^{\top})^{\top}$ denote the (pseudo-)true value of $\boldsymbol{\vartheta}$. Since our goal is to derive the asymptotic distribution of QLR_T under the null hypothesis, we assume that $\boldsymbol{\vartheta}_0 \in \boldsymbol{\Theta}_0$. Under the assumptions outlined in the next section, $\boldsymbol{\vartheta}_0$ represents the probability limit of both the restricted and unrestricted estimators defined above. We initialize the filter at $\widehat{f}_1(\boldsymbol{\vartheta}, \beta) = \omega \in \Theta_{\omega}$. This is intuitive: only in this way $\widehat{f}_t(\boldsymbol{\vartheta}_0, \beta) = \omega_0$ for all $t \in \mathbb{Z}^+$ including $t = 1$, and thus the time-varying parameter is truly constant under the null. It also renders the log-likelihood and its derivatives with respect to $(\omega, \boldsymbol{\phi})$ under the null independent from β for every sample size T , such that we can rewrite $\widehat{\mathcal{L}}_T(\boldsymbol{\vartheta}_0, \beta)$ as $\widehat{\mathcal{L}}_{0,T}(\widehat{\boldsymbol{\vartheta}}_{0,T}) = \widehat{\mathcal{L}}_T(\widehat{\boldsymbol{\vartheta}}_{0,T}, \beta)$ for any $\beta \in \mathcal{B}$, and

$$QLR_T = -2 \left(\widehat{\mathcal{L}}_{0,T}(\widehat{\boldsymbol{\vartheta}}_{0,T}) - \sup_{\beta \in \mathcal{B}} \widehat{\mathcal{L}}_T(\widehat{\boldsymbol{\vartheta}}_{\beta,T}, \beta) \right) + o_{\mathbb{P}}(1). \quad (2.5)$$

For more details on the initialization of the filter, see also [Andrews \(2001, Assumption 10 and Section 5.2 \(p. 711\)\)](#). In the next section, we study the asymptotic behavior of QLR_T for $T \rightarrow \infty$.

3 The asymptotic null distribution of the QLR test

As it is key for the formulation of the null hypothesis and the boundary problem as well as for the nonidentification problem under the null hypothesis, we first focus on the structure of the Cartesian product parameter space $\boldsymbol{\Theta} = \Theta_{\alpha} \times \Theta_{\omega} \times \boldsymbol{\Theta}_{\phi}$. With respect to the central parameter of interest α , we assume that its parameter space Θ_{α} is given by $\Theta_{\alpha} := [\alpha_L, \alpha_U]$ for some $\alpha_U \in (0, \infty)$. For α_L , we consider two scenarios: a boundary scenario with $\alpha_L = 0$, such that $\alpha_0 = 0$ lies on the boundary of Θ_{α} , and an interior point scenario with $\alpha_L \in (-\infty, 0)$, where $\alpha_0 = 0$ becomes an interior point. For ω we assume $\Theta_{\omega} = [\omega_L, \omega_U] \subset \mathbb{R}$ for some $\omega_L < \omega_U$ and $\omega_0 \in \text{int}(\Theta_{\omega})$, where $\text{int}(\cdot)$ denotes the set of interior points. This assumption

is standard in all of the score-driven time-series literature, and we adopt it throughout for notational and technical simplicity when developing the asymptotics. Finally, for ϕ we assume $\Theta_\phi = \{\phi \in \mathbb{R}^m : \phi_L \preceq \phi \preceq \phi_U\}$ for some constant vectors $\phi_L \preceq \phi_U$, where $\preceq$ denotes element-wise inequality. We allow ϕ_0 to lie on the boundary of Θ_ϕ .

To further facilitate the discussion, we use the terminology of continuous left/right (l/r) partial derivatives of order k for $k \geq 1$ from Andrews (1999, Section 3.3). For this, we consider the intersection of the parameter space above with ϵ -sized cubes, $\Theta_0^{C,\epsilon} = \Theta \cap \mathcal{C}(\vartheta_0, \epsilon)$, where $\mathcal{C}(\vartheta_0, \epsilon)$ denotes an *open* cube centered at ϑ_0 with edge length 2ϵ for any $\epsilon > 0$. The projected local neighborhood for ϕ_0 is denoted by $\Theta_{0,\phi}^{C,\epsilon} = \{\phi : (\alpha, \omega, \phi^\top)^\top \in \Theta_0^{C,\epsilon}\}$. We also define Θ_0^ϵ as a *compact* subset of $\Theta \cap \mathcal{S}(\vartheta_0, \epsilon)$, where $\mathcal{S}(\vartheta_0, \epsilon)$ denotes an *open* sphere centered at ϑ_0 with radius ϵ . Finally, we use the notation $\log^+(x) = \max\{\log(x), 0\}$ for $x > 0$, and let the p -norm of a vector $\mathbf{a} = (a_j) \in \mathbb{R}^J$ be denoted by $\|\mathbf{a}\|_p = (\sum_{j=1}^J |a_j|^p)^{1/p}$ with induced matrix norm $\|\mathbf{A}\|_p = \sup_{\mathbf{x} \neq \mathbf{0}} \|\mathbf{A}\mathbf{x}\|_p / \|\mathbf{x}\|_p$, where the subscript is omitted whenever $p = 2$. We can now formulate the following assumptions.

Assumption 1 (Model specification). *(i) Let Θ have the compact form as defined above, and let $\mathcal{B} \subset (-1, 1)$ also be compact. Furthermore, let $(\mathcal{F}, |\cdot|)$ be a complete, separable metric space with $\Theta_\omega \subseteq \mathcal{F}$.*

(ii) The set $\{(\mathbf{y}, \mathbf{x}, f, \phi) \in \mathcal{Y} \times \mathcal{X} \times \mathcal{F} \times \Theta_\phi : p(\mathbf{y} | \mathbf{x}, f, \phi) = 0\}$ has measure zero with respect to the appropriate dominating measure on $\mathcal{Y} \times \mathcal{X} \times \mathcal{F} \times \Theta_\phi$.

(iii) Both $\ell(\cdot)$ and $s(\cdot)$ are jointly measurable as functions on $\mathcal{Y} \times \mathcal{X} \times \mathcal{F} \times \Theta_\phi$ with respect to the product Borel σ -algebra, where all spaces are endowed with their Borel σ -algebras. Also assume that for every $(\mathbf{y}, \mathbf{x}) \in \mathcal{Y} \times \mathcal{X}$ the functions $(f, \phi) \mapsto \ell(\mathbf{y}, \mathbf{x}, f, \phi)$ and $(f, \phi) \mapsto s(\mathbf{y}, \mathbf{x}, f, \phi)$ are continuous on $\mathcal{F} \times \Theta_\phi$ and admit continuous l/r partial derivatives of order two with respect to (f, ϕ) on $\mathcal{F} \times \Theta_{0,\phi}^{C,\epsilon}$, where these partial derivatives are jointly measurable on $\mathcal{Y} \times \mathcal{X} \times \mathcal{F} \times \Theta_{0,\phi}^{C,\epsilon}$ for some $\epsilon > 0$.

Assumption 1 is mild in typical applications and serves as a foundational condition throughout the analysis. The requirement that the set $\{(\mathbf{y}, \mathbf{x}, f, \phi) \in \mathcal{Y} \times \mathcal{X} \times \mathcal{F} \times \Theta_\phi : p(\mathbf{y} | \mathbf{x}, f, \phi) = 0\}$ has measure zero simply ensures that the quasi-log-likelihood $\log p(\mathbf{y} | \mathbf{x}, f, \phi)$ is a.s. well defined, and in principle could be replaced by weaker conditions.

For $k \geq 1$ and $\beta \in \mathcal{B}$, define $\mathbf{f}_t^{(k)}(\vartheta, \beta) = \text{vec}(\partial \mathbf{f}_t^{(k-1)}(\vartheta, \beta) / \partial \vartheta^\top)$, where the (i, j) th entry of $\partial \mathbf{f}_t^{(k-1)}(\vartheta, \beta) / \partial \vartheta^\top$ corresponds to the l/r partial derivative of the i th element of $\mathbf{f}_t^{(k-1)}(\vartheta, \beta)$ with respect to the j th component of ϑ . We define $\mathbf{f}_t^{(0)}(\vartheta, \beta) := f_t(\vartheta, \beta)$ and, analogously, $\hat{\mathbf{f}}_t^{(0)}(\vartheta, \beta) := \hat{f}_t(\vartheta, \beta)$.

Assumption 2 (Filter invertibility). (i) $\{(\mathbf{y}_t, \mathbf{x}_t), t \in \mathbb{Z}\}$ is a strictly stationary and ergodic (SE) sequence.

(ii) Let $\mathcal{D}_k = \Theta \times \mathcal{B}$ for $k = 0$, and $\mathcal{D}_k = \Theta_0^\epsilon \times \mathcal{B}$ for $k = 1, 2$, where $\epsilon > 0$ is specified in [Assumption 1](#). For $k = 0, 1, 2$, there exist unique SE sequences $\{\mathbf{f}_t^{(k)}, t \in \mathbb{Z}\}$ and constants $\rho_k > 1$, such that $\rho_k^t \sup_{(\boldsymbol{\vartheta}, \beta) \in \mathcal{D}_k} \|\hat{\mathbf{f}}_t^{(k)}(\boldsymbol{\vartheta}, \beta) - \mathbf{f}_t^{(k)}(\boldsymbol{\vartheta}, \beta)\| \xrightarrow{a.s.} 0$, where $\mathbf{f}_t^{(k)}(\boldsymbol{\vartheta}, \beta)$ is $\mathcal{F}_{t-1}$ -measurable for every $t \in \mathbb{Z}$ and $(\boldsymbol{\vartheta}, \beta) \in \mathcal{D}_k$. Moreover, $\hat{\mathbf{f}}_t^{(k)}(\cdot)$ for $t \in \mathbb{Z}^+$ and $\mathbf{f}_t^{(k)}(\cdot)$ for $t \in \mathbb{Z}$ are uniformly continuous on $\mathcal{D}_k$.

The assumption that the variables of interest and the regressors are jointly SE is standard in the literature and may not be easily relaxed, as one typically requires the stochastic recurrence equation (SRE) in (2.1) to be SE in order to apply the results of [Bougerol \(1993\)](#) and [Straumann and Mikosch \(2006\)](#). [Assumption 2\(ii\)](#) requires the sequences $\{\hat{\mathbf{f}}_t^{(k)}(\boldsymbol{\vartheta}, \beta), t \in \mathbb{Z}^+\}$ to converge exponentially fast almost surely (e.a.s.) to a unique limiting sequence $\{\mathbf{f}_t(\boldsymbol{\vartheta}, \beta), t \in \mathbb{Z}\}$. The convergence has to be uniform over $\Theta \times \mathcal{B}$ for the filter itself ($k = 0$), and uniformly over the local parameter space $\Theta_0^\epsilon \times \mathcal{B}$ for the filter derivatives. It is also known as the uniform invertibility condition (see, e.g., [Blasques, Gorgi, Koopman, and Wintenberger, 2018](#)). The current high-level assumptions are typically satisfied for specific models from the existing score-driven time series literature as they are required for consistency and asymptotic normality of the MLE, which is typically proved in these papers. Some general results are, for instance, available in [Blasques et al. \(2022\)](#). Later in this section, we provide several flexible sufficient conditions that ensure [Assumption 2\(ii\)](#) is satisfied.

To formulate the final conditions for the asymptotic distribution of the QLR statistic, we need some notation for the log-likelihood function and its derivatives. Let $\frac{\partial}{\partial \boldsymbol{\vartheta}} \ell_t(\mathbf{f}_t(\boldsymbol{\vartheta}, \beta), \boldsymbol{\phi})$ and $\frac{\partial^2}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}^\top} \ell_t(\mathbf{f}_t(\boldsymbol{\vartheta}, \beta), \boldsymbol{\phi})$ denote the $(m+2)$ -dimensional vector and $(m+2) \times (m+2)$ matrix of l/r first and second order partial derivatives of $\ell_t(\mathbf{f}_t(\boldsymbol{\vartheta}, \beta), \boldsymbol{\phi})$ with respect to $\boldsymbol{\vartheta}$, respectively. As we want to refer to some of its sub-elements in the expression of the limiting distribution, we also define $\nabla_t^\phi(f, \boldsymbol{\phi}) := \frac{\partial}{\partial \boldsymbol{\phi}} \ell_t(f, \boldsymbol{\phi})$, $\nabla_t^{ff}(f, \boldsymbol{\phi}) := \frac{\partial}{\partial f} \nabla_t^f(f, \boldsymbol{\phi})$, $\nabla_t^{\phi f}(f, \boldsymbol{\phi}) := \frac{\partial}{\partial \boldsymbol{\phi}} \nabla_t^f(f, \boldsymbol{\phi})$, and $\nabla_t^{\phi\phi}(f, \boldsymbol{\phi}) := \frac{\partial}{\partial \boldsymbol{\phi}} (\nabla_t^\phi(f, \boldsymbol{\phi}))^\top$. We group these as $\nabla_t(f, \boldsymbol{\phi})^\top = (\nabla_t^f(f, \boldsymbol{\phi}), \nabla_t^\phi(f, \boldsymbol{\phi}))^\top$ and define $\boldsymbol{\Sigma} = \begin{pmatrix} \Sigma_{ff} & \Sigma_{f\phi}^\top \\ \Sigma_{\phi f} & \Sigma_{\phi\phi} \end{pmatrix} = \mathbb{E}[\nabla_t(\omega_0, \boldsymbol{\phi}_0) \nabla_t(\omega_0, \boldsymbol{\phi}_0)^\top \mid \mathcal{F}_{t-1}]$ and $\boldsymbol{\Omega} = \begin{pmatrix} \Omega_{ff} & \Omega_{f\phi}^\top \\ \Omega_{\phi f} & \Omega_{\phi\phi} \end{pmatrix} = -\mathbb{E}[\partial \nabla_t(\omega_0, \boldsymbol{\phi}_0) / \partial (f, \boldsymbol{\phi}^\top) \mid \mathcal{F}_{t-1}]$. With this notation in place, we now formulate the following assumption.

Assumption 3 (Asymptotic distribution of QLR_T). Let $d_t(\boldsymbol{\vartheta}, \beta) := \ell_t(\hat{\mathbf{f}}_t(\boldsymbol{\vartheta}, \beta), \boldsymbol{\phi}) - \ell_t(\mathbf{f}_t(\boldsymbol{\vartheta}, \beta), \boldsymbol{\phi})$ for $t \geq 1$ and $(\boldsymbol{\vartheta}, \beta) \in \Theta \times \mathcal{B}$.

(i) $\sup_{(\boldsymbol{\vartheta}, \beta) \in \Theta \times \mathcal{B}} |d_t(\boldsymbol{\vartheta}, \beta)| \xrightarrow{e.a.s.} 0$ as $t \rightarrow \infty$ and $\mathbb{E}[\sup_{(\boldsymbol{\vartheta}, \beta) \in \Theta \times \mathcal{B}} |\ell_t(\mathbf{f}_t(\boldsymbol{\vartheta}, \beta), \boldsymbol{\phi})|] < \infty$.

- (ii) For any $\epsilon > 0$ and $\beta \in \mathcal{B}$, $\sup_{\boldsymbol{\vartheta} \in \boldsymbol{\Theta} \setminus \mathcal{S}(\boldsymbol{\vartheta}_0, \epsilon)} \mathbb{E}[\ell_t(f_t(\boldsymbol{\vartheta}, \beta), \boldsymbol{\phi})] < \mathbb{E}[\ell_t(f_t(\boldsymbol{\vartheta}_0, \beta), \boldsymbol{\phi}_0)]$, where $\boldsymbol{\Theta} \setminus \mathcal{S}(\boldsymbol{\vartheta}_0, \epsilon)$ consists of all $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$ outside the open sphere $\mathcal{S}(\boldsymbol{\vartheta}_0, \epsilon)$.
- (iii) For $\epsilon > 0$ specified in [Assumption 1](#), $\mathbb{E}\left(\sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta}_\epsilon \times \mathcal{B}} \left\| \frac{\partial^2}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}^\top} \ell_t(f_t(\boldsymbol{\vartheta}, \beta), \boldsymbol{\phi}) \right\| \right) < \infty$.
- (iv) It holds a.s. that $\mathbb{E}[\nabla_t(\omega_0, \boldsymbol{\phi}_0) \mid \mathcal{F}_{t-1}] = \mathbf{0}$, and the matrices $\boldsymbol{\Omega}$ and $\boldsymbol{\Sigma}$ are nonrandom, finite, and positive definite for $\boldsymbol{\Omega}$, and positive semi-definite for $\boldsymbol{\Sigma}$, respectively. The lower-left block $\boldsymbol{\Omega}_{\phi f} = \mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\phi}} \nabla_t^f(\omega_0, \boldsymbol{\phi}_0) \mid \mathcal{F}_{t-1}\right]$ of $\boldsymbol{\Omega}$ equals zero, or $\mathbb{E}(h_{0,t}(\beta)) = 0$ for any $\beta \in \mathcal{B}$, or both, where $h_{0,t}(\beta) := \sum_{j=0}^{\infty} \beta^j s(\mathbf{y}_{t-j}, \mathbf{x}_{t-j}, \omega_0, \boldsymbol{\phi}_0)$.
- (v) $\mathbb{E}[s(\mathbf{y}_t, \mathbf{x}_t, \omega_0, \boldsymbol{\phi}_0)]^2 < \infty$, and $\inf_{\beta \in \mathcal{B}} \text{Var}(h_{0,t}(\beta)) > 0$.

Assumptions [3\(i\)-\(ii\)](#) imply that $\sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta} \times \mathcal{B}} |T^{-1} \widehat{\mathcal{L}}_T(\boldsymbol{\vartheta}, \beta) - \mathbb{E}[\ell_t(f_t(\boldsymbol{\vartheta}, \beta), \boldsymbol{\phi})]| \xrightarrow{a.s.} 0$ and that $\boldsymbol{\vartheta}_0$ is uniquely identifiable for every value of $\beta \in \mathcal{B}$. These assumptions are typically met by the score-driven models in the literature. For instance, it is often not hard to verify $\sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta} \times \mathcal{B}} |d_t(\boldsymbol{\vartheta}, \beta)| \xrightarrow{e.a.s.} 0$ in [Assumption 3\(i\)](#). Using a mean value theorem for scalar functions, one obtains

$$\sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta} \times \mathcal{B}} |d_t(\boldsymbol{\vartheta}, \beta)| \leq \sup_{(f, \boldsymbol{\vartheta}) \in \mathcal{F} \times \boldsymbol{\Theta}} |\nabla_t^f(f, \boldsymbol{\phi})| \sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta} \times \mathcal{B}} |\hat{f}_t(\boldsymbol{\vartheta}, \beta) - f_t(\boldsymbol{\vartheta}, \beta)|. \quad (3.1)$$

For robust score-driven filters such as the Student's t location model ([Harvey and Luati, 2014](#)), $\sup_{(f, \boldsymbol{\vartheta}) \in \mathcal{F} \times \boldsymbol{\Theta}} |\nabla_t^f(f, \boldsymbol{\phi})|$ is bounded by construction. Then the result immediately follows from the filter invertibility in [Assumption 2\(ii\)](#) for $k = 0$. More generally, when $\sup_{(f, \boldsymbol{\vartheta}) \in \mathcal{F} \times \boldsymbol{\Theta}} |\nabla_t^f(f, \boldsymbol{\phi})|$ is not bounded, but identically distributed and satisfies $\mathbb{E}(\log^+ \sup_{(f, \boldsymbol{\vartheta}) \in \mathcal{F} \times \boldsymbol{\Theta}} |\nabla_t^f(f, \boldsymbol{\phi})|) < \infty$, one also has $\sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta} \times \mathcal{B}} |d_t(\boldsymbol{\vartheta}, \beta)| \xrightarrow{e.a.s.} 0$ by invoking Lemma 2.1 of [Straumann and Mikosch \(2006\)](#) along with [Assumption 2\(ii\)](#) for $k = 0$.

[Assumption 3\(iii\)](#) requires identifiable uniqueness of the model under the null hypothesis. An important special case is that of correct specification under the null of constant parameters. If $p(\mathbf{y} \mid \mathbf{x}, f, \boldsymbol{\phi}) = p(\mathbf{y} \mid \mathbf{x}, \tilde{f}, \tilde{\boldsymbol{\phi}})$ if and only if $f = \tilde{f}$ and $\boldsymbol{\phi} = \tilde{\boldsymbol{\phi}}$ for almost every $(\mathbf{y}, \mathbf{x}) \in \mathcal{Y} \times \mathcal{X}$ (with respect to an appropriate dominating measure on $\mathcal{Y} \times \mathcal{X}$), then [Assumption 3\(ii\)](#) holds; see [Online Appendix A](#) for a justification. In this case, we thus only require identification of the dynamic parameter f_t and the static parameter $\boldsymbol{\phi}$ characterizing the pmf or pdf. This is easily satisfied in many cases, such as for instance location and scale models (see [Creal et al., 2013](#); [Harvey and Luati, 2014](#)).

[Assumption 3\(iii\)](#) implies that, for some $\epsilon > 0$,

$$\sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta}_\epsilon \times \mathcal{B}} \left\| T^{-1} \frac{\partial^2}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}^\top} \mathcal{L}_T(\boldsymbol{\vartheta}, \beta) - \mathbb{E} \left(\frac{\partial^2}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}^\top} \ell_t(f_t(\boldsymbol{\vartheta}, \beta), \boldsymbol{\phi}) \right) \right\| \xrightarrow{a.s.} 0. \quad (3.2)$$

Verifying this condition typically requires [Assumption 2\(ii\)](#) for $k = 1, 2$, i.e., the invertibility of the (perturbed) SREs for the derivative processes of $\hat{f}_t(\boldsymbol{\vartheta}, \beta)$ with respect to $\boldsymbol{\vartheta}$; see the results in [Proposition 1](#) below.

Finally, Assumptions [3\(iv\)–\(v\)](#) impose moment conditions that are needed to derive a limiting approximation of QLR_T . Specifically, they require that (1) $\left\{ \left(\nabla_t^f(\omega_0, \boldsymbol{\phi}_0), \nabla_t^\phi(\omega_0, \boldsymbol{\phi}_0)^\top \right)^\top, t \in \mathbb{Z} \right\}$ forms a martingale difference sequence (m.d.s.) with respect to $\{\mathcal{F}_t, t \in \mathbb{Z}\}$, such that we can apply a central limit theorem for m.d.s.; (2) $\mathbb{E} \left(\frac{\partial^2}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}^\top} \ell_t(f_t(\boldsymbol{\vartheta}, \beta), \boldsymbol{\phi}) \Big|_{\boldsymbol{\vartheta}=\boldsymbol{\vartheta}_0} \right)$ has a block-diagonal structure, ensuring that either $(\hat{\alpha}_{\beta,T}, \hat{\omega}_{\beta,T})^\top$ and $\hat{\boldsymbol{\phi}}_{\beta,T}$, or $\hat{\alpha}_{\beta,T}$ and $(\hat{\omega}_{\beta,T}, \hat{\boldsymbol{\phi}}_{\beta,T}^\top)^\top$, can be separately approximated asymptotically; and (3) the asymptotic variance of QLR_T exists and is non-zero. Again, as we see later in several examples, these conditions are typically verified in score-driven models available in the existing literature. The m.d.s. assumption in (1) above is not as restrictive as it may appear at first sight. If the model is mis-specified, we have to interpret ω_0 and $\boldsymbol{\phi}_0$ as the pseudo-true values under the null hypothesis, which ensures that the m.d.s. property is satisfied by design even in these cases.

For some $p \in \mathbb{Z}^+$, let $\Rightarrow$ denote weak convergence in the space of $\mathbb{R}^p$ -valued continuous functions on $\mathcal{B}$, equipped with the uniform metric, following [Pollard \(1990\)](#). We then have the following main result.

Theorem 1 (Limiting null distribution of QLR_T). *If Assumptions 1–3 hold, then*

$$\left\{ \frac{1}{\sqrt{T}} \frac{\partial}{\partial \boldsymbol{\vartheta}} \mathcal{L}_T(\boldsymbol{\vartheta}, \beta) \Big|_{\boldsymbol{\vartheta}=\boldsymbol{\vartheta}_0} \right\}_{\beta \in \mathcal{B}} \Rightarrow \{\mathbf{G}(\beta)\}_{\beta \in \mathcal{B}}, \quad T \rightarrow \infty, \quad (3.3)$$

where $\{\mathbf{G}(\beta)\}_{\beta \in \mathcal{B}}$ is a zero-mean Gaussian process with $\mathbf{G}(\beta) \in \mathbb{R}^{m+2}$ and covariance function

$$\text{Cov}(\mathbf{G}(\beta_1), \mathbf{G}(\beta_2)) = \begin{pmatrix} \Sigma_{ff} \mathbb{E}[h_{0,t}(\beta_2) h_{0,t}(\beta_1)] & \Sigma_{ff} \bar{h}_0(\beta_1) & \boldsymbol{\Sigma}_{\phi f}^\top \bar{h}_0(\beta_1) \\ \Sigma_{ff} \bar{h}_0(\beta_2) & \Sigma_{ff} & \boldsymbol{\Sigma}_{\phi f}^\top \\ \boldsymbol{\Sigma}_{\phi f} \bar{h}_0(\beta_2) & \boldsymbol{\Sigma}_{\phi f} & \boldsymbol{\Sigma}_{\phi \phi} \end{pmatrix}, \quad (3.4)$$

where $\bar{h}_0(\beta) = \mathbb{E}[h_{0,t}(\beta)]$ with $h_{0,t}(\beta) = \sum_{j=0}^{\infty} \beta^j s(\mathbf{y}_{t-j}, \mathbf{x}_{t-j}, \omega_0, \boldsymbol{\phi}_0)$ as defined in [Assumption 3](#).

In addition,

$$QLR_T \xrightarrow{d} \sup_{\beta \in \mathcal{B}} \left\{ \Phi(\beta)^{-1} [Z(\beta)]^2 \right\}, \quad T \rightarrow \infty, \quad (3.5)$$

where we distinguish the following cases for $\Phi(\beta)$ and $Z(\beta)$:

(i) if $\boldsymbol{\Omega}_{\phi f} = \mathbf{0}$ in [Assumption 3\(iv\)](#), then $\Phi(\beta) := \mathbf{e}_1^\top \mathcal{J}_{(\alpha, \omega)}^{-1}(\beta) \mathbf{e}_1$, where $\mathbf{e}_1 = (1, 0)^\top$, and

$$\mathcal{J}_{(\alpha, \omega)}(\beta) := \Omega_{ff} \begin{pmatrix} \mathbb{E}[h_{0,t}^2(\beta)] & \bar{h}_0(\beta) \\ \bar{h}_0(\beta) & 1 \end{pmatrix}, \quad (3.6)$$

while $Z(\beta) := \max \{ \mathbf{e}_1^\top \mathcal{J}_{(\alpha, \omega)}^{-1}(\beta) \mathbf{G}_{(\alpha, \omega)}(\beta), 0 \}$ if $\alpha_L = 0$, and $Z(\beta) := \mathbf{e}_1^\top \mathcal{J}_{(\alpha, \omega)}^{-1}(\beta) \mathbf{G}_{(\alpha, \omega)}(\beta)$ if $\alpha_L < 0$, with $\mathbf{G}_{(\alpha, \omega)}(\beta)$ denoting the first two components of $\mathbf{G}(\beta)$.

(ii) if in [Assumption 3\(iv\)](#) $\boldsymbol{\Omega}_{\phi f}$ is possibly nonzero, but $\bar{h}_0(\beta) = 0$, then $\Phi(\beta)^{-1} := \mathcal{J}_\alpha(\beta) = \Omega_{ff} \mathbb{E}(h_{0,t}^2(\beta))$, while $Z(\beta) := \max \{ \mathcal{J}_\alpha^{-1}(\beta) G_\alpha(\beta), 0 \}$ if $\alpha_L = 0$, and $Z(\beta) := \mathcal{J}_\alpha^{-1}(\beta) G_\alpha(\beta)$ if $\alpha_L < 0$, with $G_\alpha(\beta)$ denoting the first component of $\mathbf{G}(\beta)$.

The proof of [Theorem 1](#) can, in principle, accommodate any objective function $\hat{\mathcal{L}}_T$ and corresponding score s_t that satisfy Assumptions 1–3, and is therefore not limited to a likelihood-based framework. As such, [Theorem 1](#) may also extend to recent variants such as the quasi score-driven models proposed by [Blasques, Francq, and Laurent \(2023\)](#). The limiting distribution in (3.5) takes a non-standard form and has a different distribution whether we only have the nonidentification problem under H_0 (case $\alpha_L < 0$) or also the boundary problem (case $\alpha_L = 0$). Note that in cases where the parameter ϕ does not enter the conditional pdf/pmf, i.e., $p(\cdot \mid \mathbf{x}_t, f_t, \phi) = p(\cdot \mid \mathbf{x}_t, f_t)$, all assumptions involving ϕ in Assumptions 1–3 can be skipped and the result simplifies further. An example of this is the well-known class of Poisson autoregressive models (i.e., integer GARCH) of [Fokianos et al. \(2009\)](#); see Section 4.3. Also in these cases, however, the asymptotic distribution of the QLR_T , remains non-standard.

[Theorem 1](#) distinguishes two cases. Case (i) with $\boldsymbol{\Omega}_{\phi f} = \mathbf{0}$ is also found in [Andrews \(2001\)](#). This condition, however, is not always easily satisfied for score-driven models. For instance, in a setting with a time-varying scale for a Student's t distribution, the degrees of freedom parameter is typically correlated with the scale parameter. Therefore, Case (ii) in [Theorem 1](#) generalizes the result to the setting with nonzero $\boldsymbol{\Omega}_{\phi f}$.

The limiting distribution in (3.5) generally depends on nuisance parameters and may therefore be difficult to operationalize in practice. We discuss two corollaries highlighting special cases where the asymptotic distribution is free of such nuisance parameters and that may be better suited for inference. We first formulate the following assumption.

Assumption 4. (i) $\{s(\mathbf{y}_t, \mathbf{x}_t, \omega_0, \phi_0), t \in \mathbb{Z}\}$ forms a m.d.s. with respect to the filtration $\{\mathcal{F}_t, t \in \mathbb{Z}\}$. (ii) There exists an estimator $\hat{\kappa}_{G,T}$ such that $\hat{\kappa}_{G,T} = \Omega_{ff}^{-1} \Sigma_{ff} + o_{\mathbb{P}}(1)$, where $o_{\mathbb{P}}(1)$ is uniform in $\beta \in \mathcal{B}$.

An important case where [Assumption 4\(i\)](#) is satisfied is when $S(\mathbf{y}_t, \mathbf{x}_t, \omega_0, \phi_0)$ is $\mathcal{F}_{t-1}$ -measurable. This is true for virtually all score-driven models in the literature, where the scaling matrix is typically taken as a power of the inverse conditional Fisher information matrix following the suggestions in [Creal et al. \(2013\)](#). In this case, under the null-hypothesis of static parameters, the score has expectation zero as the (pseudo)true parameters ω_0 and ϕ_0 satisfy the first-order condition for the limit log-likelihood objective function, minimizing the Kullback–Leibler divergence between the possibly mis-specified conditional pdf/pmf $p(\cdot \mid \mathbf{x}_t, f_t, \phi)$ and the unobserved true model. We have the following result.

Corollary 1. *If Assumptions 1–4 hold, the scaled QLR test statistic $\widetilde{QLR}_T := \hat{\kappa}_{G,T}^{-1} QLR_T$ converges in distribution to a limit that is free of nuisance parameters:*

$$\widetilde{QLR}_T \xrightarrow{d} \begin{cases} \sup_{\beta \in \mathcal{B}} \left(\max \left\{ (1 - \beta^2)^{1/2} \sum_{j=0}^{\infty} \beta^j X_j, 0 \right\} \right)^2, & \alpha_L = 0, \\ \sup_{\beta \in \mathcal{B}} \left((1 - \beta^2)^{1/2} \sum_{j=0}^{\infty} \beta^j X_j \right)^2, & \alpha_L < 0, \end{cases} \quad (3.7)$$

where $\{X_j, j \geq 0\}$ is a sequence of i.i.d. standard normal random variables and $\hat{\kappa}_{G,T}$ is defined in [Assumption 4\(ii\)](#).

The limiting distribution in (3.7) can be simulated in a straightforward way to obtain critical values and conduct inference. [Table B.3](#) in Online Appendix B provides a comprehensive set of values. It is worth mentioning that the process $\{(1 - \beta^2)^{1/2} \sum_{j=0}^{\infty} \beta^j X_j\}_{\beta \in \mathcal{B}}$ is well documented in the literature, having been found, for instance, by [Andrews \(2001, Section 5.2\)](#) for tests of conditional heteroskedasticity and by [Andrews and Ploberger \(1996, Theorem 1\)](#) for tests of serial correlation in ARMA(1,1) models. To operationalize the scaled $\widetilde{QLR}_T$ statistic, we can take $\hat{\kappa}_{G,T} = \left(-T^{-1} \sum_{t=1}^T \nabla_t^{ff}(\hat{\omega}_{0,T}, \hat{\phi}_{0,T}) \right)^{-1} \left(T^{-1} \sum_{t=1}^T [\nabla_t^f(\hat{\omega}_{0,T}, \hat{\phi}_{0,T})]^2 \right)$, which, under specific conditions, satisfies [Assumption 4\(ii\)](#) under the null hypothesis. However, since $f_t(\boldsymbol{\vartheta}, \beta)$ no longer reduces to its unconditional mean under H_1 , such an estimator $\hat{\kappa}_{G,T}$ may perform badly under the alternative, which may have a negative impact on the power of $\widetilde{QLR}_T$. We can therefore also consider the alternatively scaled test statistic $\widetilde{QLR}_T^\dagger := -2 \left(\sup_{\beta \in \mathcal{B}} \left(\hat{\mathcal{L}}_{0,T}(\hat{\boldsymbol{\vartheta}}_{0,T}) / \hat{\kappa}_{G,T}(\beta) \right) - \sup_{\beta \in \mathcal{B}} \left(\hat{\mathcal{L}}_T(\hat{\boldsymbol{\vartheta}}_{\beta,T}, \beta) / \hat{\kappa}_{G,T}(\beta) \right) \right) + o_{\mathbb{P}}(1)$, with $\hat{\kappa}_{G,T}(\beta) = \left(-T^{-1} \sum_{t=1}^T \nabla_t^{ff}(\hat{f}_t(\hat{\boldsymbol{\vartheta}}_{\beta,T}, \beta), \hat{\phi}_{\beta,T}) \right)^{-1} \left(T^{-1} \sum_{t=1}^T [\nabla_t^f(\hat{f}_t(\hat{\boldsymbol{\vartheta}}_{\beta,T}, \beta), \hat{\phi}_{\beta,T})]^2 \right)$.

Finally, the following special case is of particular interest.

Corollary 2. *If Assumptions 1–3 and [Assumption 4\(i\)](#) hold, and if $\Sigma_{ff} = \Omega_{ff}$, then QLR_T shares the same limiting distribution as $\widetilde{QLR}_T$ as given in Eq. (3.7).*

Note that the additional condition in [Corollary 2](#) holds, for instance, if one is willing to assume that the static version of the model is correctly specified. In that case, the information matrix equality will automatically hold for the parameter f in the pdf/pmf expression $p(\cdot \mid \mathbf{x}_t, f_t, \phi)$.

As already mentioned, for many score-driven models sufficient conditions for uniform filter invertibility over the entire parameter space $\Theta \times \mathcal{B}$ have been established, thus ensuring the conditions in [Assumption 2](#). These sufficient conditions typically build on [Bougerol \(1993\)](#) and [Straumann and Mikosch \(2006\)](#); see for instance [Blasques et al. \(2022\)](#). For completeness, sufficient conditions for the case $k = 0$ are presented in [Lemma A.1](#) in the appendix. To conclude this section, we provide some new conditions for the invertibility and the existence of moments of the *derivative* processes of the filter with respect to the static parameters. The existence of such moments is (implicitly) assumed by the moment conditions for the derivatives of the log-likelihood contributions in, for example, [Assumption 3\(iii\)](#). The key novelty in the next set of conditions is to exploit the fact that moments conditions only need to hold in a small ball around the null parameter space, which we indicated by $\Theta_0^\epsilon \times \mathcal{B}$. This can substantially relax earlier sufficient conditions formulated over the entire parameter space and may therefore prove useful in practical settings for specific models.

We rewrite partial derivatives using subscripts, e.g., $s_\phi(\mathbf{y}_t, \mathbf{x}_t, f, \phi) = \partial s(\mathbf{y}_t, \mathbf{x}_t, f, \phi) / \partial \phi$, $s_{ff}(\mathbf{y}_t, \mathbf{x}_t, f, \phi) = \partial^2 s(\mathbf{y}_t, \mathbf{x}_t, f, \phi) / \partial f^2$, $s_{(f,\phi)(f,\phi)}(\mathbf{y}_t, \mathbf{x}_t, f, \phi) = \partial s(\mathbf{y}_t, \mathbf{x}_t, f, \phi) / \partial \left(\begin{smallmatrix} f \\ \phi \end{smallmatrix} \right) \partial (f, \phi^\top)$. We also define the shorthand notation $\bar{s}_t^\epsilon = \sup_{(\boldsymbol{\vartheta}, \beta) \in \Theta_0^\epsilon \times \mathcal{B}} |s(\mathbf{y}_t, \mathbf{x}_t, f_t(\boldsymbol{\vartheta}, \beta), \phi)|$, and similar notation for l/r derivatives, e.g., $\bar{s}_{\phi,t}^\epsilon = \sup_{(\boldsymbol{\vartheta}, \beta) \in \Theta_0^\epsilon \times \mathcal{B}} \|s_\phi(\mathbf{y}_t, \mathbf{x}_t, f, \phi)|_{f=f_t(\boldsymbol{\vartheta}, \beta)}\|$ and $\bar{s}_{(f,\phi)(f,\phi),t}^\epsilon = \sup_{\boldsymbol{\vartheta} \in \Theta_0^\epsilon} \|s_{(f,\phi)(f,\phi)}(\mathbf{y}_t, \mathbf{x}_t, f, \phi)|_{f=f_t(\boldsymbol{\vartheta}, \beta)}\|$, where we make explicit that the supremum is only taken over an ϵ -ball of $\boldsymbol{\vartheta}_0$. We also define the double supremum $\bar{\bar{s}}_{f,t}^\epsilon = \sup_{\boldsymbol{\vartheta} \in \Theta_0^\epsilon} \sup_{f \in \mathcal{F}} |\beta + \alpha s_f(\mathbf{y}_t, \mathbf{x}_t, f, \phi)|$. We can now formulate the following assumption.

Assumption 5 (Invertibility of derivative processes). *Let $\eta, \kappa \in [1, \infty]$ satisfy $\eta^{-1} + \kappa^{-1} \leq 1$, and let $\epsilon > 0$ be the constant specified in [Assumption 1](#).*

- (i) *If for all $t \in \mathbb{Z}$ there is an $M < \infty$ such that $\bar{s}_t^\epsilon + \bar{s}_{\phi,t}^\epsilon \leq M$ almost surely, then set $\eta = \infty$. Otherwise, for $\eta < \infty$, it holds that $\mathbb{E}[(\bar{s}_t^\epsilon)^\eta + (\bar{s}_{\phi,t}^\epsilon)^\eta] < \infty$. Moreover, $\mathbb{E}[\log \bar{\bar{s}}_{f,t}^\epsilon] < 0$.*
- (ii) *If $s_{f,t}(\boldsymbol{\vartheta}, \beta)$ is nonrandom, set $\kappa = \infty$ and require*

$$\sum_{j=1}^{\infty} \sup_{(\boldsymbol{\vartheta}, \beta) \in \Theta_0^\epsilon \times \mathcal{B}} \left| \prod_{\ell=1}^j \left(\beta + \alpha \frac{\partial s(\mathbf{y}_{t-\ell}, \mathbf{x}_{t-\ell}, f_{t-\ell}(\boldsymbol{\vartheta}, \beta), \phi)}{\partial f} \right) \right| < \infty.$$

Otherwise, for $\kappa < \infty$ and for every $j \in \mathbb{Z}^+$, there exists a positive real sequence $\{\varrho_j(\boldsymbol{\vartheta}_0, \epsilon, \kappa), j \geq$

1} with $\sum_{j=1}^{\infty} \varrho_j(\boldsymbol{\vartheta}_0, \epsilon, \kappa)^{1/\kappa} < \infty$ and

$$\mathbb{E} \left[\sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta}_0^\epsilon \times \mathcal{B}} \left| \prod_{\ell=1}^j \left(\beta + \alpha \frac{\partial s(\mathbf{y}_{t-\ell}, \mathbf{x}_{t-\ell}, f_{t-\ell}(\boldsymbol{\vartheta}, \beta), \phi)}{\partial f} \right) \right|^\kappa \right] \leq \varrho_j(\boldsymbol{\vartheta}_0, \epsilon, \kappa), \quad \forall t \in \mathbb{Z}.$$

(iii) For all $t \in \mathbb{Z}$ and $f, f^\dagger \in \mathcal{F}$, the following Lipschitz conditions hold:

$$\sup_{\boldsymbol{\vartheta} \in \boldsymbol{\Theta}_0^\epsilon} |s(\mathbf{y}_t, \mathbf{x}_t, f, \phi) - s(\mathbf{y}_t, \mathbf{x}_t, f^\dagger, \phi)| \leq C_t(\boldsymbol{\vartheta}_0, \epsilon) |f - f^\dagger|^\zeta, \quad (3.8)$$

$$\sup_{\boldsymbol{\vartheta} \in \boldsymbol{\Theta}_0^\epsilon} \|s_{(f, \phi)}(\mathbf{y}_t, \mathbf{x}_t, f, \phi) - s_{(f, \phi)}(\mathbf{y}_t, \mathbf{x}_t, f^\dagger, \phi)\| \leq C_t(\boldsymbol{\vartheta}_0, \epsilon) |f - f^\dagger|^\zeta, \quad (3.9)$$

$$\sup_{\boldsymbol{\vartheta} \in \boldsymbol{\Theta}_0^\epsilon} \|s_{(f, \phi)(f, \phi)}(\mathbf{y}_t, \mathbf{x}_t, f, \phi) - s_{(f, \phi)(f, \phi)}(\mathbf{y}_t, \mathbf{x}_t, f^\dagger, \phi)\| \leq C_t(\boldsymbol{\vartheta}_0, \epsilon) |f - f^\dagger|^\zeta, \quad (3.10)$$

for some constant $\zeta \in (0, 1]$ and a strictly stationary scalar sequence $\{C_t(\boldsymbol{\vartheta}_0, \epsilon), t \in \mathbb{Z}\}$ with $\mathbb{E}[\log^+ C_t(\boldsymbol{\vartheta}_0, \epsilon)] < \infty$.

(iv) If $\bar{s}_{(f, \phi), t}^\epsilon \leq M$ and $\bar{s}_{(f, \phi)(f, \phi), t}^\epsilon \leq M$ for all $t \in \mathbb{Z}$ and some $M < \infty$, i.e., if the relevant partial derivatives of the score are almost surely uniformly bounded, then $2\eta^{-1} + 3\kappa^{-1} \leq 1$. If this uniform boundedness condition does not hold and either $\eta < \infty$ or $\kappa < \infty$ (or both), then there exists a $\lambda \geq 2$ satisfying $2\lambda^{-1} + \kappa^{-1} \leq 1$ and $\lambda^{-1} \geq 2(\eta^{-1} + \kappa^{-1})$, such that $\mathbb{E}[|\bar{s}_{f, t}^\epsilon|^\lambda] < \infty$, $\mathbb{E}[\|\bar{s}_{ff, t}^\epsilon\|^\lambda] < \infty$, $\mathbb{E}[\|\bar{s}_{f\phi, t}^\epsilon\|^\lambda] < \infty$, and $\mathbb{E}[\|\bar{s}_{\phi\phi, t}^\epsilon\|^{\lambda/2}] < \infty$. If both $\eta = \kappa = \infty$, these last four moment conditions also hold, but for some arbitrary $\lambda \geq 2$.

Assumption 5(i) implies the existence of the first-order moment of the score process and its first-order derivative. Although we do not explicitly require the derivative processes to have at least a first-order moment, such a moment typically arises when verifying **Assumption 3(i)**. This moment condition is also standard in the literature when establishing the asymptotic normality of static parameter estimators. For robust filters, such as the Student's t location model of [Harvey and Luati \(2014\)](#), **Assumption 5(i)** generally holds with $\eta = \infty$. The contraction condition $\mathbb{E}[\log \bar{s}_{f, t}^\epsilon] < 0$ in **Assumption 5(i)** implies the local invertibility of the filter and builds on standard conditions of [Bougerol \(1993\)](#). It is typically satisfied for score-driven models, which often require uniform invertibility of the filter for consistency of parameter estimates. The contraction condition can be further relaxed by considering multiple iterations of the stochastic recurrence equation (SRE); see [Lemma A.1](#) in [Online Appendix A](#). **Assumption 5(ii)** is similar to, but weaker than, Assumption AN3 in [Lin and Lucas \(2025\)](#). It is needed to ensure the existence of appropriate moments for the partial derivative processes of the filter with respect to $\boldsymbol{\vartheta}$. This condition is relatively straightforward to verify, particularly when $s_t(f, \phi)$ is linear

in f . For instance, in the Gaussian location model, where $\partial s_{t-\ell}(f, \phi)/\partial f = -1$, choosing $\varrho_j(\boldsymbol{\vartheta}_0, \epsilon, \kappa) = \sup_{(\boldsymbol{\vartheta}, \beta) \in \Theta_0^\epsilon \times \mathcal{B}} |\beta - \alpha|^{j\kappa}$ reduces the condition to requiring $\sup_{(\boldsymbol{\vartheta}, \beta) \in \Theta_0^\epsilon \times \mathcal{B}} |\beta - \alpha| < 1$. [Assumption 5\(iii\)](#) imposes *local* Lipschitz continuity conditions around $\boldsymbol{\vartheta}_0$ on the forcing variable $s_t(\cdot)$ and its partial derivatives up to second order. Similarly, [Assumption 5\(iv\)](#), like [Assumption 5\(i\)](#), implies that also the second-order derivative process has at least a first moment, which is useful for verifying [Assumption 3\(iii\)](#). We summarize this in the following result.

Proposition 1 (Invertibility of derivative processes). *Let $\epsilon > 0$ denote the constant specified in [Assumption 1](#). Let [Assumption 2\(ii\)](#) hold for $k = 0$, and let [Assumption 5](#) also hold, then [Assumption 2\(ii\)](#) also holds for $k = 1, 2$, where $\mathbb{E}\left(\sup_{(\boldsymbol{\vartheta}, \beta) \in \Theta_0^\epsilon \times \mathcal{B}} \|\mathbf{f}_t^{(1)}(\boldsymbol{\vartheta}, \beta)\|^{\delta_1}\right) < \infty$ for $\delta_1 = (\eta^{-1} + \kappa^{-1})^{-1}$ if $\eta < \infty$ or $\kappa < \infty$, and for any $\delta_1 \geq 1$ if $\eta = \kappa = \infty$. Moreover, $\mathbb{E}\left(\sup_{(\boldsymbol{\vartheta}, \beta) \in \Theta_0^\epsilon \times \mathcal{B}} \|\mathbf{f}_t^{(2)}(\boldsymbol{\vartheta}, \beta)\|^{\delta_2}\right) < \infty$ for $\delta_2 = (2\delta_1^{-1} + \kappa^{-1})^{-1}$ if a uniform bound M applies in [Assumption 5\(iv\)](#), and for $\delta_2 = (2\lambda^{-1} + \kappa^{-1})^{-1}$ otherwise.*

4 Examples

To illustrate the theory developed thus far and show how the conditions formulated in [Assumptions 1–5](#) can be applied, we study four different models. [Sections 4.1 and 4.2](#) consider a correctly specified Gaussian location and Student’s t volatility model, respectively. [Section 4.3](#) analyzes a Poisson autoregressive model for discrete time series, allowing for possible mis-specification. As all three models are univariate and exclude exogenous variables, [Section 4.4](#) investigates a multivariate spatial model with random regressors. We outline only the key steps in the main text; full details are provided in [Online Appendix C](#).

4.1 Gaussian location model

For $t = 1, \dots, T$, consider the model $y_t = f_t + u_t$, where $u_t = \sigma_u \epsilon_t$ and ϵ_t are i.i.d. standard normal random variables. Assume that the sequence $\{y_t, t \in \mathbb{Z}\}$ is generated under the null hypothesis, i.e., $y_t \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(\omega_0, \sigma_{u,0}^2)$. Under this assumption, the model is correctly specified. Note that $\nabla_t^f(f, \phi) = (y_t - f)/\sigma_u^2$, where $\phi = \sigma_u > 0$. Take $S_t(f, \phi) = \sigma_u^2$, also known as inverse information matrix scaling (see [Creal et al., 2013](#)), then $s_t(f, \phi) = y_t - f$ and $f_{t+1} = \omega(1 - \beta) + \beta f_t + \alpha(y_t - f_t)$. We define $\mathcal{B} = [0, \beta_U]$ with $\beta_U < 1$, and $\Theta = \{\boldsymbol{\vartheta} = (\alpha, \omega, \phi)^\top \in \mathbb{R}^3 : 0 \leq \alpha \leq \alpha_U < 1, \omega_L \leq \omega \leq \omega_U, 0 < \sigma_L \leq \phi \leq \sigma_U\}$. Let $\mathcal{F} = \mathbb{R}$, so that [Assumption 1](#) trivially holds for any $\epsilon > 0$.

[Assumption 2\(i\)](#) and [Assumption 2\(ii\)](#) for $k = 0$ follow directly by standard arguments

from Bougerol (1993) using $\mathbb{E} \log^+ |y_t| < \infty$, $\omega_U < \infty$, and $\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} |\beta - \alpha| = \alpha_U \vee \beta_U < 1$. To verify Assumption 2(ii) for $k = 1, 2$, we apply Proposition 1, which in turn requires checking Assumption 5. Assumption 5(i) holds for arbitrary $\eta \geq 1$, such as $\eta = 4$. Assumption 5(ii) is fulfilled with $\kappa = \infty$, because $\sum_{j=1}^{\infty} \sup_{(\boldsymbol{\theta}, \beta) \in \boldsymbol{\Theta}_0^\epsilon \times \mathcal{B}} \left| \prod_{\ell=1}^j (\beta - \alpha) \right| \leq \sum_{j=1}^{\infty} (\alpha_U \vee \beta_U)^j < \infty$. Assumption 5(iii) holds with $C_t(\boldsymbol{\theta}_0, \epsilon) = 1$. Assumption 5(iv) holds given all the relevant suprema are either 1 or 0. Assumption 2(ii) for $k = 1, 2$ then follows immediately from Proposition 1.

For Assumption 3(i), note that

$$\begin{aligned} \sup_{(\boldsymbol{\theta}, \beta) \in \boldsymbol{\Theta} \times \mathcal{B}} |d_t(\boldsymbol{\theta}, \beta)| &\leq \frac{1}{2} \sigma_L^{-2} \left(\sup_{(\boldsymbol{\theta}, \beta) \in \boldsymbol{\Theta} \times \mathcal{B}} |\hat{f}_t(\boldsymbol{\theta}, \beta) - f_t(\boldsymbol{\theta}, \beta)| \right)^2 \\ &\quad + \sigma_L^{-2} \sup_{(\boldsymbol{\theta}, \beta) \in \boldsymbol{\Theta} \times \mathcal{B}} |y_t - f_t(\boldsymbol{\theta}, \beta)| \sup_{(\boldsymbol{\theta}, \beta) \in \boldsymbol{\Theta} \times \mathcal{B}} |\hat{f}_t(\boldsymbol{\theta}, \beta) - f_t(\boldsymbol{\theta}, \beta)|. \end{aligned}$$

By Assumption 2 and Lemma 2.1 of Straumann and Mikosch (2006), $\sup_{(\boldsymbol{\theta}, \beta) \in \boldsymbol{\Theta} \times \mathcal{B}} |d_t(\boldsymbol{\theta}, \beta)| \xrightarrow{e.a.s.} 0$ provided that $\mathbb{E}[\log^+ \sup_{(\boldsymbol{\theta}, \beta) \in \boldsymbol{\Theta} \times \mathcal{B}} |y_t - f_t(\boldsymbol{\theta}, \beta)|] < \infty$, which follows immediately from $\mathbb{E}[\sup_{(\boldsymbol{\theta}, \beta) \in \boldsymbol{\Theta} \times \mathcal{B}} |y_t - f_t(\boldsymbol{\theta}, \beta)|^4] < \infty$, which also ensures that $\mathbb{E}[\sup_{(\boldsymbol{\theta}, \beta) \in \boldsymbol{\Theta} \times \mathcal{B}} |\ell_t(f_t(\boldsymbol{\theta}, \beta), \boldsymbol{\phi})|] < \infty$. Assumption 3(iii) can be obtained using the Cauchy–Schwarz inequality, together with $\mathbb{E}\left(\sup_{(\boldsymbol{\theta}, \beta) \in \boldsymbol{\Theta}_0^\epsilon \times \mathcal{B}} \|\mathbf{f}_t^{(1)}(\boldsymbol{\theta}, \beta)\|^4\right) < \infty$ and $\mathbb{E}\left(\sup_{(\boldsymbol{\theta}, \beta) \in \boldsymbol{\Theta}_0^\epsilon \times \mathcal{B}} \|\mathbf{f}_t^{(2)}(\boldsymbol{\theta}, \beta)\|^2\right) < \infty$. Assumption 3(iv) holds immediately with $\boldsymbol{\Sigma} = \boldsymbol{\Omega} = \text{diag}(\sigma_{u,0}^{-2}, 2\sigma_{u,0}^{-2})$. Since the $s_t(\omega_0, \boldsymbol{\phi}_0) = u_t$ are independent, the result $QLR_T \xrightarrow{d} \sup_{\beta \in \mathcal{B}} \left(\max \left\{ (1 - \beta^2)^{1/2} \sum_{j=0}^{\infty} \beta^j X_j, 0 \right\} \right)^2$ follows directly by applying Corollary 2.

4.2 t -GAS volatility

In our second example, we consider the time-varying scale model of Creal et al. (2013), also examined in Harvey (2013) and Blasques et al. (2022), to illustrate how the theory can be applied to non-linear filters. The model is given by $y_t = f_t^{1/2} u_t$, where u_t are i.i.d. t_ν random variables with $\nu > 0$ degrees of freedom. Here we can also illustrate how Proposition 1 results in weaker assumptions than those found in the literature. Let $\boldsymbol{\phi} = \nu$ and define the score-induced observation weight $w_t(f, \boldsymbol{\phi}) = (1 + \nu^{-1}) / (1 + \nu^{-1} y_t^2 / f)$. Then $\nabla_t^f(f, \boldsymbol{\phi}) = 2^{-1} f^{-2} (w_t(f, \boldsymbol{\phi}) \cdot y_t^2 - f)$. Taking $S_t(f, \boldsymbol{\phi}) = 2f^2$ as in Creal et al. (2013), we obtain $s_t(f, \boldsymbol{\phi}) = w_t(f, \boldsymbol{\phi}) \cdot y_t^2 - f$. Let $\mathcal{F} = \mathbb{R}^+$ and define $\mathcal{B} = [\beta_L, \beta_U] \subset (0, 1)$. Consider $\boldsymbol{\Theta} = \{\boldsymbol{\theta} = (\alpha, \omega, \boldsymbol{\phi})^\top \in \mathbb{R}^3 : 0 \leq \alpha \leq \beta_L, 0 < \omega_L \leq \omega \leq \omega_U, 0 < \nu_L \leq \boldsymbol{\phi} \leq \nu_U\}$. This specification guarantees the positivity of the filter. Furthermore, the filter is bounded from below, as $f_t \geq \frac{\omega_L(1 - \beta_U)}{1 - \beta_U + \beta_L} \forall t \in \mathbb{Z}$.

Assumption 1 holds for any $\epsilon > 0$. Note that $\{y_t, t \in \mathbb{Z}\}$ is an SE sequence and that Assumption 2(ii) for $k = 0$ holds following the arguments in, for instance, Blasques et al. (2022).

To verify [Assumption 2\(ii\)](#) for $k = 1, 2$, we again check the conditions in [Proposition 1](#). Note that $\sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta}_0^\epsilon \times \mathcal{B}} |s_t(f_t(\boldsymbol{\vartheta}, \beta), \boldsymbol{\phi})| \leq C \sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta}_0^\epsilon \times \mathcal{B}} |f_t(\boldsymbol{\vartheta}, \beta)|$, where $\sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta}_0^\epsilon \times \mathcal{B}} |f_t(\boldsymbol{\vartheta}, \beta)|$ is bounded because

$$\begin{aligned} f_{t+1}(\boldsymbol{\vartheta}, \beta) &= \omega(1 - \beta) + (\beta - \alpha)f_t(\boldsymbol{\vartheta}, \beta) + \alpha \frac{1 + \nu^{-1}}{1 + \nu^{-1}y_t^2/f_t(\boldsymbol{\vartheta}, \beta)} y_t^2 \\ &\leq \omega(1 - \beta) + (\beta + \alpha\nu)f_t(\boldsymbol{\vartheta}, \beta) \leq \omega_U(1 - \beta_L) + (\beta_U + \epsilon\nu_U)f_t(\boldsymbol{\vartheta}, \beta). \end{aligned}$$

[Assumption 1](#) allows us to pick a sufficiently small $\epsilon > 0$ with $\beta_U + \epsilon\nu_U < 1$, ensuring that $\sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta}_0^\epsilon \times \mathcal{B}} |f_t(\boldsymbol{\vartheta}, \beta)|$ is bounded. Compared to the condition $\sup_{\boldsymbol{\vartheta} \in \boldsymbol{\Theta}} (\beta + \nu\alpha) < 1$ as imposed by [Blasques et al. \(2022, p. 331\)](#), we do not require such a strong restriction on the parameter space as [Assumption 5](#) only considers a local neighborhood $\boldsymbol{\Theta}_0^\epsilon$ of $\boldsymbol{\vartheta}_0$. Similarly, it is straightforward to show that $\bar{s}_{\boldsymbol{\phi}, t}^\epsilon$ is bounded. Hence, [Assumption 5\(i\)](#) holds for $\eta = \infty$. [Assumption 5\(ii\)](#) holds for any $\kappa \geq 1$ and some sufficiently small $\epsilon > 0$ if we set $\varrho_j(\boldsymbol{\vartheta}_0, \epsilon, \kappa) = (\beta_U + \epsilon\nu_U)^{j\kappa}$ with $\beta_U + \epsilon\nu_U < 1$. Note that if we first compute the l/r partial derivatives of the mapping $(f, \boldsymbol{\phi}) \mapsto s_t(f, \boldsymbol{\phi})$ and then substitute $f = f_t(\boldsymbol{\vartheta}, \beta)$, the resulting derivatives are bounded (locally) up to at least third order for all $(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta}_0^\epsilon \times \mathcal{B}$. Using a mean value theorem (e.g., [Rudin, 1976, Theorem 9.19](#)), [Assumption 5\(iii\)](#) is then satisfied using $\zeta = 1$ and $C_t(\boldsymbol{\vartheta}_0, \epsilon) = C$ for some $C > 0$ for all $t \in \mathbb{Z}$. Moreover, the second derivatives of the score can be shown to be uniformly bounded, such that [Assumption 5\(iv\)](#) holds. Then [Proposition 1](#) holds, where δ_1 and δ_2 are allowed to take arbitrarily large values.

For [Assumption 3\(i\)](#), we have $\sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta} \times \mathcal{B}} |d_t(\boldsymbol{\vartheta}, \beta)| \xrightarrow{e.a.s.} 0$ by applying the invertibility results from [Blasques et al. \(2022\)](#). Moreover, $\mathbb{E}[\sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta} \times \mathcal{B}} |\ell_t(f_t(\boldsymbol{\vartheta}, \beta), \boldsymbol{\phi})|] < \infty$ follows directly from the boundedness of $f_t(\boldsymbol{\vartheta}, \beta)$ and the existence of a logarithmic moment for y_t . [Assumption 3\(iii\)](#) holds by applying [Proposition 1](#), using the form of the log-likelihood for the Student's t distribution. By somewhat tedious linear algebra, together with the distributional assumption on y_t and the results in [Lin and Lucas \(2025, Eq. \(E.6\)\)](#) and [Harvey \(2013, Proposition 39, p. 211\)](#), we obtain $\Sigma_{ff} = \Omega_{ff} = \nu_0/[2\omega_0^2(\nu_0 + 3)]$ and $\boldsymbol{\Omega}_{\boldsymbol{\phi}f} = [\omega_0(\nu_0 + 3)(\nu_0 + 1)]^{-1} > 0$, i.e., the scale and the degrees of freedom estimators are correlated in general. However, since $\mathbb{E}(s_t(\omega_0, \boldsymbol{\phi}_0) | \mathcal{F}_{t-1}) = 0$, it follows that $\mathbb{E}(h_{0,t}(\beta)) = 0$ for any $\beta \in \mathcal{B}$. All remaining conditions in [Assumptions 3\(iv\)–\(v\)](#) and [Corollary 2](#) are checked easily. Therefore, QLR_T converges to the limiting distribution given in [\(3.7\)](#).

4.3 Poisson autoregression

Our third example illustrates how the theory works out for integer data and for the case of mis-specified models. For this, we take the example of a Poisson time series model $y_t \mid \mathcal{F}_{t-1} \sim \text{Poisson}(f_t)$, where $f_t > 0$ for every $t \in \mathbb{Z}$. In this case, there is no static parameter ϕ , so all conditions involving ϕ and the corresponding components in [Theorem 1](#) can be disregarded, as noted below [Theorem 1](#). Moreover, in this example we allow for model mis-specification. That is, the statistician observes the count data $\{y_t, t = 1, \dots, T\}$, which need not follow a Poisson distribution in the dgp. We nevertheless require that $\{y_t, t \in \mathbb{Z}\}$ is an SE sequence with $\mathbb{E}|y_t|^4 < \infty$, $\mathbb{E}(y_t \mid \mathcal{F}_{t-1}) = \omega_0$, and $\mathbb{E}((y_t - \omega_0)^2 \mid \mathcal{F}_{t-1}) = \text{Var}(y_t)$. Since $\nabla_t^f(f, \phi) = (y_t - f)/f$, we obtain $f_{t+1} = \omega(1 - \beta) + \beta f_t + \alpha(y_t - f_t)$ by setting $S_t(f, \phi) = f$. This yields the well-known integer GARCH model of [Fokianos et al. \(2009\)](#). In what follows, we verify the conditions for [Theorem 1](#). As the verification closely parallels the Gaussian location example in [Section 4.1](#), we focus only on the key differences.

Let $\mathcal{F} = \mathbb{R}^+$ and define $\mathcal{B} = [\beta_L, \beta_U] \subset (0, 1)$. Set $\Theta = \{\boldsymbol{\vartheta} = (\alpha, \omega)^\top \in \mathbb{R}^2 : 0 \leq \alpha \leq \beta_L, 0 < \omega_L \leq \omega \leq \omega_U\}$. The specifications of Θ and $\mathcal{B}$ ensure the positivity of f_t . Thus, [Assumption 1](#) holds. Furthermore, [Assumption 2](#) holds since the filter is invertible (see [Fokianos et al., 2009](#)) and [Proposition 1](#) applies with $\delta_1 = 4$ and $\delta_2 = 2$. [Assumption 3\(i\)](#) is then also directly satisfied. To verify [Assumption 3\(ii\)](#), we apply Jensen's inequality to the function $f \mapsto \omega_0 \log(f) - f$ for $f > 0$ and $\omega_0 > 0$, and obtain

$$\begin{aligned} \mathbb{E}\left(\ell_t(f_t(\boldsymbol{\vartheta}, \beta), \phi)\right) &= \mathbb{E}\left(\omega_0 \log(f_t(\boldsymbol{\vartheta}, \beta)) - f_t(\boldsymbol{\vartheta}, \beta)\right) - \mathbb{E}(\log(y_t!)) \\ &\leq \left(\omega_0 \log\left(\mathbb{E}(f_t(\boldsymbol{\vartheta}, \beta))\right) - \mathbb{E}(f_t(\boldsymbol{\vartheta}, \beta))\right) - \mathbb{E}(\log(y_t!)), \end{aligned}$$

where equality holds if and only if $f_t(\boldsymbol{\vartheta}, \beta) = \mathbb{E}(f_t(\boldsymbol{\vartheta}, \beta)) > 0$ a.s., that is, $f_t(\boldsymbol{\vartheta}, \beta)$ is a.s. constant. Since $s_t(f, \phi) = y_t - f_t$ is random, one must have $\alpha = 0 = \alpha_0$, which implies $f_t(\boldsymbol{\vartheta}, \beta) = \omega$ for any $\beta \in \mathcal{B}$ a.s. As $\mathbb{E}(\ell_t(f_t(\boldsymbol{\vartheta}, \beta), \phi)) \leq \omega_0 \log(\omega) - \omega - \mathbb{E}(\log(y_t!)) \leq \omega_0 \log(\omega_0) - \omega_0 - \mathbb{E}(\log(y_t!))$, with equality if and only if $\omega = \omega_0$, we thus verified [Assumption 3\(ii\)](#). Similarly, [Assumption 3\(iii\)–\(iv\)](#) hold with $\Sigma_{ff} = \text{Var}(y_t)/\omega_0^2 < \infty$ and $\Omega_{ff} = -\mathbb{E}(\nabla_t^{ff}(\omega_0, \phi_0)) = \omega_0^{-1} \in (0, \infty)$. Finally, we have $\inf_{\beta \in \mathcal{B}} \text{Var}\left(\sum_{j=0}^{\infty} \beta^j s_t(\omega_0, \phi_0)\right) = (1 - \beta_L^2)^{-1} \text{Var}(y_t) > 0$. Therefore, [Assumption 3\(v\)](#) is fulfilled. By [Corollary 1](#), (3.7) holds for $\widetilde{QLR}_T$ provided that $\hat{\kappa}_{G,n} = \text{Var}(y_t)/\omega_0 + o_{\mathbb{P}}(1)$. If, in addition, $\text{Var}(y_t) = \omega_0$, then QLR_T converges to the same limiting distribution as $\widetilde{QLR}_T$ by [Corollary 2](#).

4.4 Spatial dynamic spillovers

Our final example illustrates the theory for multivariate, non-linear models with additional exogenous variables. We consider the spatial regression model of Blasques et al. (2016): $\mathbf{y}_t = \tilde{\rho}(f_t) \mathbf{W} \mathbf{y}_t + \mathbf{x}_t \boldsymbol{\gamma} + \mathbf{e}_t$, where $\mathbf{W}$ is an $n \times n$ matrix of exogenous, nonrandom spatial weights and spectral radius smaller than 1, $\mathbf{x}_t$ is (with a slight abuse of notation) a $n \times k$ matrix of regressors, $\boldsymbol{\gamma}$ is an $k \times 1$ vector of static parameters, and $\mathbf{e}_t$ is an $n \times 1$ disturbance vector. Here, both n and k are finite. The link function is specified as $\tilde{\rho}(\cdot) := \rho \tanh(\cdot)$, where $\rho \in (0, 1)$ is a user-specified parameter. Note that $\tilde{\rho}(f) \in (-\rho, \rho)$ for all $f \in \mathcal{F}$, and that its derivatives of every order are bounded. Suppose the statistician specifies that $\mathbf{e}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{V})$ are i.i.d., where $\mathbf{V} = \text{diag}(\sigma_1^2, \dots, \sigma_n^2)$ is positive definite. Let $\boldsymbol{\phi} = (\boldsymbol{\gamma}^\top, \sigma_1, \dots, \sigma_n)^\top$. We have $\ell_t(f, \boldsymbol{\phi}) = -2^{-1}n \log(2\pi) + \log \det(\mathbf{I}_n - \tilde{\rho}(f)\mathbf{W}) - 2^{-1} \log \det(\mathbf{V}) - 2^{-1}(\mathbf{y}_t - \tilde{\rho}(f)\mathbf{W}\mathbf{y}_t - \mathbf{x}_t\boldsymbol{\gamma})^\top \mathbf{V}^{-1}(\mathbf{y}_t - \tilde{\rho}(f)\mathbf{W}\mathbf{y}_t - \mathbf{x}_t\boldsymbol{\gamma})$. As in Blasques et al. (2016), we set $S_t(f, \boldsymbol{\phi}) = 1$. Then the forcing variable $s_t(f, \boldsymbol{\phi})$ in (2.1) is given by $s_t(f, \boldsymbol{\phi}) = \nabla_t^f(f, \boldsymbol{\phi}) = \left(\mathbf{y}_t^\top \mathbf{W}^\top \mathbf{V}^{-1} [\mathbf{y}_t - \tilde{\rho}(f)\mathbf{W}\mathbf{y}_t - \mathbf{x}_t\boldsymbol{\gamma}] - \text{tr}(\mathbf{Z}(f)\mathbf{W}) \right) \dot{\tilde{\rho}}(f)$, where $\mathbf{Z}(f) := (\mathbf{I}_n - \tilde{\rho}(f)\mathbf{W})^{-1}$ and $\dot{\tilde{\rho}}(f) = \rho \cdot (1 - \tanh^2(f))$.

To study the null distribution of QLR_T under misspecification, we assume that $\mathbf{y}_t$ is generated as $\mathbf{y}_t = \tilde{\rho}(\omega_0)\mathbf{W}\mathbf{y}_t + \mathbf{x}_t\boldsymbol{\gamma}_0 + \boldsymbol{\varepsilon}_t$, where the error term $\boldsymbol{\varepsilon}_t := (\varepsilon_{1,t}, \dots, \varepsilon_{n,t})^\top$ has zero mean, and $\mathbb{E}[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top | \mathbf{x}_t, \mathcal{F}_{t-1}] = \mathbf{V}_0$, with $\mathbf{V}_0 = \text{diag}(\sigma_{0,1}^2, \dots, \sigma_{0,n}^2)$ diagonal and positive definite. However, we do not require $\boldsymbol{\varepsilon}_t$ to follow the same distribution as $\mathbf{e}_t$. We require that $\{(\mathbf{y}_t, \mathbf{x}_t), t \in \mathbb{Z}\}$ is SE with $\mathbb{E}[\|\boldsymbol{\varepsilon}_t\|^4 + \|\mathbf{x}_t\|^4] < \infty$, $\mathbb{E}[\mathbf{x}_t^\top \mathbf{x}_t]$ is positive definite, and $\mathbb{E}[\mathbf{y}_t^\top \mathbf{W}^\top \mathbf{V}_0^{-1} \boldsymbol{\varepsilon}_t - \text{tr}(\mathbf{Z}(\omega_0)\mathbf{W})]^2 > 0$. Moreover, $\mathbb{E}(\boldsymbol{\varepsilon}_t | \mathbf{x}_t, \mathcal{F}_{t-1}) = \mathbf{0}$ a.s., and $\mathbb{E}[\boldsymbol{\varepsilon}_t^\top \otimes (\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top) | \mathbf{x}_t, \mathcal{F}_{t-1}]$, $\mathbb{E}[(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top) \otimes (\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top) | \mathbf{x}_t, \mathcal{F}_{t-1}]$, $\mathbb{E}[\mathbf{x}_t | \mathcal{F}_{t-1}]$, and $\mathbb{E}[\mathbf{x}_t^\top \otimes \mathbf{x}_t | \mathcal{F}_{t-1}]$ are all nonrandom and bounded a.s., where $\otimes$ denotes the Kronecker product. We impose a technical assumption that there exist $1 \leq k < i \leq n$ such that $\sigma_{0,i}^2[\mathbf{W}\mathbf{Z}(\omega_0)]_{ki} + \sigma_{0,k}^2[\mathbf{W}\mathbf{Z}(\omega_0)]_{ik} \neq 0$, where $[\mathbf{W}\mathbf{Z}(\omega_0)]_{ki}$ denotes the (k, i) th element of $\mathbf{W}\mathbf{Z}(\omega_0)$.

To verify the different assumptions formulated in Section 3, we set $\mathcal{F} = \mathbb{R}$, $\mathcal{B} = [\beta_L, \beta_U] \subset (-1, 1)$, and $\boldsymbol{\Theta} = \{\boldsymbol{\vartheta} = (\alpha, \omega, \boldsymbol{\gamma}^\top, \sigma_1, \dots, \sigma_n)^\top \in \mathbb{R}^{k+n+2} : \alpha_L \leq \alpha \leq \alpha_U, \omega_L \leq \omega \leq \omega_U, \boldsymbol{\gamma}_L \preceq \boldsymbol{\gamma} \preceq \boldsymbol{\gamma}_U, 0 < \sigma_L \leq \sigma_i \leq \sigma_U, i = 1, \dots, n\}$, where $\alpha_L < 0$. Verifying Assumption 2(ii) requires tedious linear algebra. For details, we refer interested readers to Blasques et al. (2016). Here, for brevity, we simply assume these conditions are satisfied, with $\mathbb{E}\left(\sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta}_0^\epsilon \times \mathcal{B}} \|\mathbf{f}_t^{(1)}(\boldsymbol{\vartheta}, \beta)\|^4\right) < \infty$ and $\mathbb{E}\left(\sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta}_0^\epsilon \times \mathcal{B}} \|\mathbf{f}_t^{(2)}(\boldsymbol{\vartheta}, \beta)\|^2\right) < \infty$ for some $\epsilon > 0$ and that identification condition Assumption 3(ii) holds. The detailed steps in Online Appendix C.4 then show that all remaining conditions in Assumptions 1–3 are satisfied. For instance, since $\mathbb{E}(\|\boldsymbol{\varepsilon}_t\|^4 + \|\mathbf{x}_t\|^4) < \infty$, it follows that $\mathbb{E}(\|\mathbf{y}_t\|^4) \leq 8\|\mathbf{Z}(\omega_0)\|^4 \mathbb{E}(\|\mathbf{x}_t\|^4 \|\boldsymbol{\gamma}_0\|^4 + \|\boldsymbol{\varepsilon}_t\|^4) < \infty$,

and $\mathbb{E}\left(\sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta}_0^\epsilon \times \mathcal{B}} \|\mathbf{e}_t(f_t(\boldsymbol{\vartheta}, \beta), \boldsymbol{\gamma})\|^4\right) < \infty$ for some $\epsilon > 0$. By repeatedly applying the Cauchy-Schwarz inequality, together with the above moment conditions and the quantities computed below, we can then establish [Assumption 3\(iii\)](#). Similarly, as we have $\mathbb{E}[\nabla_t^f(\omega_0, \boldsymbol{\phi}_0) | \mathbf{x}_t, \mathcal{F}_{t-1}] = \left[\text{tr}\left(\mathbf{Z}(\omega_0)^\top \mathbf{W}^\top \mathbf{V}_0^{-1} \mathbb{E}_{t-1}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top)\right) - \text{tr}(\mathbf{Z}(\omega_0) \mathbf{W}) \right] \dot{h}(\omega_0) = 0$, we immediately obtain $\mathbb{E}(h_{0,t}(\beta)) = 0$ for any $\beta \in \mathcal{B}$, thus verifying part of [Assumption 3\(iv\)](#). Detailed checks of all remaining assumptions can be found in Online Appendix C.4 and illustrate that the theory also readily applies to these more complex models.

Since [Assumption 4\(i\)](#) holds, then if $\Omega_{ff}^{-1} \Sigma_{ff}$ can be estimated consistently using a plug-in estimator $\hat{\Omega}_{ff}^{-1} \hat{\Sigma}_{ff}$, [Corollary 1](#) yields $\widetilde{QLR}_T \xrightarrow{d} \sup_{\beta \in \mathcal{B}} \left((1 - \beta^2)^{1/2} \sum_{j=0}^{\infty} \beta^j X_j \right)^2$. If the model is correctly specified, $\Omega_{ff}^{-1} \Sigma_{ff} = 1$ and [Corollary 2](#) then shows also QLR_T converges to this limiting distribution.

5 Simulations

We examine the empirical size and power of our QLR test across the four different models of Section 4. We consider correctly specified models with the following parameter settings. First, for the Gaussian location model in Section 4.1, we set $(\omega_0, \sigma_{u0}) = (1, 1)$, $\beta_U = \alpha_U = 0.95$. Since $\alpha_L = \beta_L = 0$, according to [Table B.3](#) in the Online Appendix B, this yields critical values of 2.989, 4.308, and 7.277 for nominal levels of 10%, 5%, and 1%, respectively. Second, for the t -GAS volatility model in Section 4.2, we set $(\omega_0, \nu_0) = (2, 4)$ such that we have fat, non-Gaussian tails, and $(\beta_L, \beta_U) = (0.3, 0.95)$, which give critical values of 2.816, 4.104, and 7.055 for $\alpha_L = 0$. Third, for the Poisson autoregressive model in Section 4.3, we set $\omega_0 = 10$ and $(\beta_L, \beta_U) = (0.3, 0.95)$, with critical values identical to those of the t -GAS model. Finally, for the spatial model in Section 4.4, we take $(\omega_0, \rho, n) = (1, 0.9, 3)$ and let $\mathbf{e}_t \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(\mathbf{0}_n, \sigma_{e,0}^2 \mathbf{I}_n)$ with $\sigma_{e,0} = 0.5$. To reduce computational cost, we restrict the model to have a scalar covariance matrix $\boldsymbol{\Sigma} = \sigma_e^2 \mathbf{I}_n$ with a common variance parameter σ_e^2 . We include a single regressor $\mathbf{x}_t \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(\boldsymbol{\mu}, \mathbf{I}_n)$, where $\boldsymbol{\mu} = (1, 1, \dots, 1)^\top$, with slope coefficient $\boldsymbol{\gamma}_0 = 0.2$. The spatial weight matrix $\mathbf{W}$ is randomly generated, with its diagonal elements set to zero and rows subsequently normalized, yielding $\mathbf{W} = \begin{pmatrix} 0 & 0.53 & 0.47 \\ 0.63 & 0 & 0.37 \\ 0.61 & 0.39 & 0 \end{pmatrix}$ in our simulations. We set $(\beta_L, \beta_U) = (0, 0.95)$ and $(\alpha_L, \alpha_U) = (-0.3, 0.3)$, leading to critical values of 4.247, 5.587, and 8.731 at the 10%, 5%, and 1% nominal levels, respectively. For convenience, the critical values used in this section is summarized in [Table 1](#).

We consider sample sizes $T \in \{1000, 3000\}$, with all results based on 20,000 Monte Carlo replications. To evaluate the empirical size and power of the proposed QLR test, we vary $\alpha_0 \in \{0, 0.02, 0.04, 0.06\}$, where $\alpha_0 = 0$ corresponds to empirical size and $\alpha_0 > 0$ to empirical

Table 1: Subset of simulated critical values for the limiting distribution in Eq. (3.7) used in Section 5, with $\Theta_\alpha = [\alpha_L, \alpha_U]$ for $\alpha_U > 0$ and $\mathcal{B} = [\beta_L, \beta_U]$.

β_L	β_U	$\alpha_L = 0$			$\alpha_L < 0$		
		10%	5%	1%	10%	5%	1%
0	0.950	2.989	4.308	7.277	4.247	5.587	8.731
0.3	0.950	2.816	4.104	7.055	—	—	—

Table 2: Empirical null rejection rates of QLR_T : $\alpha_0 = 0$ corresponds to the empirical size, and $\alpha_0 > 0$ to the empirical power. Values are expressed in percentages.

α_0	$T = 1000$			$T = 3000$			$T = 1000$			$T = 3000$		
	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
Gaussian location							Poisson autoregression					
0	7.85	3.72	0.76	8.88	4.16	0.79	7.48	3.76	0.67	8.13	3.95	0.84
0.02	41.60	29.43	12.81	78.70	68.00	45.39	43.26	30.74	13.74	79.32	68.82	46.30
0.04	83.15	74.43	54.40	99.64	99.26	97.23	84.80	76.04	56.35	99.78	99.47	97.72
0.06	97.69	95.58	88.08	100	100	99.99	97.98	96.01	89.22	100	100	100
t -GAS volatility							spatial spillovers					
0	7.61	3.66	0.70	8.18	3.96	0.77	11.78	6.2	1.51	10.61	5.56	1.22
0.02	24.03	14.46	4.59	45.40	32.23	13.93	99.14	98.61	96.49	100	100	100
0.04	49.11	36.33	17.56	86.44	77.89	57.55	99.99	99.99	99.99	100	100	100
0.06	73.14	61.92	40.10	98.48	96.83	90.55	100	100	100	100	100	100

power. In all four cases, when evaluating empirical power, we set $\beta = 0.9$.

The results are reported in Table 2. We see that the limiting distribution in (3.7) provides a good approximation across all four models for a relatively small sample size of $T = 1000$. In general, the QLR test tends to be slightly conservative, with the exception of the spatial model. When T increases to 3000, the empirical size approaches the corresponding nominal levels further. We also see that the empirical power rises with both larger values of α_0 and larger sample sizes. The sharp increase in empirical power observed in the spatial model compared to the other examples can be attributed to the multivariate nature of $\mathbf{y}_t$ in the spatial example, compared to the univariate nature of $\mathbf{y}_t$ in the other models.

6 Empirical application

As an illustration, we consider a simplified version of the dynamic spatial regression model studied in Blasques et al. (2016) and D’Innocenzo et al. (2024); see also the example in Section 4.4. We consider dynamic spatial credit risk spillovers between $n = 7$ European sovereigns (Germany, France, Ireland, Italy, the Netherlands, Portugal, and Spain) using weekly changes in 5-year CDS spreads over the ten-year period January 4, 2013, to December 16, 2022 such that $T = 520$. As

in D’Innocenzo et al. (2024), the regressors include an intercept, changes in short-term (Eonia) interest rates and in option-implied stock market volatility (VSTOXX), and country-specific stock index returns and changes in the long (10Y) – short (1Y) government yield spreads. The spatial weights are taken from the BIS website and relate domestic banks’ cross-border exposures and row-normalized; see D’Innocenzo et al. (2024).

Given the high persistence in the spatial spillover patterns, we set the parameter bounds to $(\beta_L, \beta_U) = (0, 0.995)$ and $(\alpha_L, \alpha_U) = (-0.3, 0.3)$, yielding a corresponding critical value of 9.127 at the 1% significance level according to Table B.3.

We estimate a static Gaussian version of the model (under H_0) and a corresponding model with a dynamic spatial spillover parameter ρ_t as in Section 4.4. The parameter estimates and log-likelihood values for both models are presented in Appendix D. The estimation results shows that the unrestricted estimate of α is very close to zero with $|\alpha| < 0.001$. Still, the resulting QLR test statistic is $QLR_T \approx 41.0$, which under standard asymptotics would be suggestive of strong statistical significance. The scaled $\widetilde{QLR}_T$ test statistic, however, is much smaller at $\widetilde{QLR}_T \approx 9.8$, where we computed $\hat{\kappa}_{G,T}$ as $\left(-T^{-1} \sum_{t=1}^T \nabla_t^{ff}(\hat{\omega}_{0,T}, \hat{\phi}_{0,T}) \right)^{-1} \left(T^{-1} \sum_{t=1}^T [\nabla_t^f(\hat{\omega}_{0,T}, \hat{\phi}_{0,T})]^2 \right)$. This scaled $\widetilde{QLR}_T$ statistic is about a factor 4 smaller than the original QLR_T statistic and only marginally significant at the 1% level. For the case of a Student’s t assumption for the error term, the differences are considerably smaller, with $QLR_T \approx 14.720$ and $\widetilde{QLR}_T \approx 16.683$, both being statistically significant against the non-standard critical value. This illustrates that directly comparing likelihoods and using standard asymptotic theory to conduct inference on the relevance of score-driven time-varying parameters can be tricky in empirically relevant situations, and that the corrections derived in Section 3 can lead to material differences in such settings.

7 Conclusion

In this paper, we proposed a quasi-likelihood ratio (QLR) test for parameter constancy against the alternative of scalar score-driven dynamics. The flexible set-up accommodated different types of variables of interest (discrete and continuous, univariate and multivariate) and time-varying parameters. In developing the asymptotic null distribution of the QLR test, we faced two main challenges. First, parameters could lie on the boundary of the parameter space. Second, some parameters could not be identified under the null hypothesis. As a result, conventional asymptotic analyses could not be applied. We found that the limiting distribution generally depended on multiple nuisance parameters. However, under mild conditions satisfied by virtually all score-driven models, a simple but non-standard distribution could be derived. The limiting

distribution is free of nuisance parameters, enabling critical values to be easily simulated. We also provided a set of easily verifiable conditions and illustrated their application in different examples involving univariate and multivariate data, correct and incorrect specification, and different types of time-varying parameters (location, scale, spatial correlation). Simulation results showed a satisfactory finite sample performance of the proposed test. To further improve finite sample accuracy, future work can explore bootstrap-based approximations next to the current asymptotic approximation. Moreover, it would be interesting to extend the methodology to test for parameter constancy in settings with more time-varying parameters and multivariate score-driven dynamics.

References

- Andrews, D. W. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica*, 821–856.
- Andrews, D. W. (1999). Estimation when a parameter is on a boundary. *Econometrica* 67(6), 1341–1383.
- Andrews, D. W. (2001). Testing when a parameter is on the boundary of the maintained hypothesis. *Econometrica* 69(3), 683–734.
- Andrews, D. W. and W. Ploberger (1994). Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica*, 1383–1414.
- Andrews, D. W. and W. Ploberger (1996). Testing for serial correlation against an ARMA(1,1) process. *Journal of the American Statistical Association* 91(435), 1331–1342.
- Armiliotta, M., P. Gorgi, and A. Lucas (2025). Copula tensor count autoregressions for modeling multidimensional integer-valued time series. Technical Report TI 2025-004/III, Tinbergen Institute Discussion Paper.
- Babii, A., X. Chen, and E. Ghysels (2019). Commercial and residential mortgage defaults: Spatial dependence with frailty. *Journal of Econometrics* 212(1), 47–77.
- Baek, Y. I., J. S. Cho, and P. C. Phillips (2015). Testing linearity using power transforms of regressors. *Journal of Econometrics* 187(1), 376–384.
- Beutner, E. A., Y. Lin, and A. Lucas (2023). Consistency, distributional convergence, and optimality of score-driven filters. Technical Report TI 2023-051/III, Tinbergen Institute Discussion Paper.
- Billingsley, P. (1999). *Convergence of Probability Measures*. John Wiley & Sons.
- Blasques, F., C. Francq, and S. Laurent (2023). Quasi score-driven models. *Journal of Econometrics* 234(1), 251–275.

- Blasques, F., P. Gorgi, S. J. Koopman, and O. Wintenberger (2018). Feasible invertibility conditions and maximum likelihood estimation for observation-driven models. *Electronic Journal of Statistics* 12(1), 1019–1052.
- Blasques, F., S. J. Koopman, and A. Lucas. (2015). Information theoretic optimality of observation driven time series models for continuous responses. *Biometrika* 102, 325–343.
- Blasques, F., S. J. Koopman, and A. Lucas. (2018). Amendments and corrections: Information theoretic optimality of observation driven time series models for continuous responses. *Biometrika* 105, 753.
- Blasques, F., S. J. Koopman, A. Lucas, and J. Schaumburg (2016). Spillover dynamics for systemic risk measurement using spatial financial time series models. *Journal of Econometrics* 195(2), 211–223.
- Blasques, F., A. Lucas, and E. Silde (2018). A stochastic recurrence equations approach for score driven correlation models. *Econometric Reviews* 37(2), 166–181.
- Blasques, F., J. van Brummelen, S. J. Koopman, and A. Lucas (2022). Maximum likelihood estimation for score-driven models. *Journal of Econometrics* 227(2), 325–346.
- Bougerol, P. (1993). Kalman filtering with random coefficients and contractions. *SIAM Journal on Control and Optimization* 31(4), 942–959.
- Calvori, F., D. Creal, S. J. Koopman, and A. Lucas (2017). Testing for parameter instability across different modeling frameworks. *Journal of Financial Econometrics* 15(2), 223–246.
- Cavaliere, G., H. B. Nielsen, R. S. Pedersen, and A. Rahbek (2022). Bootstrap inference on the boundary of the parameter space, with application to conditional volatility models. *Journal of Econometrics* 227(1), 241–263.
- Cavaliere, G., H. B. Nielsen, and A. Rahbek (2017). On the consistency of bootstrap testing for a parameter on the boundary of the parameter space. *Journal of Time Series Analysis* 38(4), 513–534.
- Cavaliere, G., I. Perera, and A. Rahbek (2024). Specification tests for GARCH processes with nuisance parameters on the boundary. *Journal of Business & Economic Statistics* 42(1), 197–214.
- Creal, D., S. J. Koopman, and A. Lucas. (2011). A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations. *Journal of Business & Economic Statistics* 29, 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.
- Davidson, J. (1994). *Stochastic Limit Theory*. Oxford: Oxford University Press.
- Davies, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* 64, 247–254.
- Davies, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* 74, 33–43.
- De Punder, R., T. Dimitriadis, and R.-J. Lange (2024). Kullback-Leibler-based characterizations of score-driven updates. *arXiv preprint arXiv:2408.02391*.

- D’Innocenzo, E., A. Lucas, A. Opschoor, and X. Zhang (2024). Heterogeneity and dynamics in network models. *Journal of Applied Econometrics* 39(1), 150–173.
- D’Innocenzo, E., A. Lucas, B. Schwaab, and X. Zhang (2024). Modeling extreme events: Time-varying extreme tail shape. *Journal of Business & Economic Statistics* 42(3), 903–917.
- Fokianos, K., A. Rahbek, and D. Tjøstheim (2009). Poisson autoregression. *Journal of the American Statistical Association* 104(488), 1430–1439.
- Francq, C. and J.-M. Zakoïan (2009). Testing the nullity of GARCH coefficients: correction of the standard tests and relative efficiency comparisons. *Journal of the American Statistical Association* 104(485), 313–324.
- Gorgi, P. (2020). Beta-negative binomial auto-regressions for modelling integer-valued time series with extreme observations. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 82(5), 1325–1347.
- Gorgi, P., C. S. Lauria, and A. Luati (2024). On the optimality of score-driven models. *Biometrika* 111(3), 865–880.
- Hansen, B. E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica*, 413–430.
- Harvey, A. C. (2013). *Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series*. Econometric Series Monographs. Cambridge: Cambridge University Press.
- 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, F., D. Li, and K. Zhu (2020). Non-standard inference for augmented double autoregressive models with null volatility coefficients. *Journal of Econometrics* 215(1), 165–183.
- Kopylev, L. and B. Sinha (2011). On the asymptotic distribution of likelihood ratio test when parameters lie on the boundary. *Sankhya B* 73(1), 20–41.
- Lin, Y. and A. Lucas (2025). Functional location-scale models with robust observation-driven dynamics. Technical Report TI 2025-027/III, Tinbergen Institute Discussion Paper.
- Lin, Y., A. Lucas, and S. Ye (2025). Matrix-valued spatial autoregressions with dynamic and robust heterogeneous spillovers. Technical Report TI 2025-042/III, Tinbergen Institute Discussion Paper.
- Massacci, D. (2017). Tail risk dynamics in stock returns: Links to the macroeconomy and global markets connectedness. *Management Science* 63(9), 3072–3089.
- Opschoor, A., P. Janus, A. Lucas, and D. Van Dijk (2018). New HEAVY models for fat-tailed realized covariances and returns. *Journal of Business & Economic Statistics* 36(4), 643–657.
- Pollard, D. (1990). Empirical Processes: Theory and Applications. In *NSF-CBMS Regional Conference Series in Probability and Statistics*, Volume 2.
- Rudin, W. (1976). *Principles of Mathematical Analysis* (3rd ed.). New York: McGraw-Hill.

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

White, H. (1996). *Estimation, Inference and Specification Analysis*. Cambridge University Press.

White, H. (2001). *Asymptotic Theory for Econometricians* (Rev. ed.). Bingley: Emerald.

A Proof of Theorem 1

To prove the theorem, we apply Theorem 4 of Andrews (2001). Under the maintained assumption that $\boldsymbol{\vartheta}_0 \in \boldsymbol{\Theta}_0$ and by applying Theorem 3 of Andrews (2001), the relevant sufficient conditions to verify are Assumptions 1*, 2*, 3*, 5*, and 7–10 of Andrews (2001). Note that the weak convergence (3.3) arises as a byproduct in the verification of Assumption 3*.

Verification of Assumption 1*. Under Assumption 3(ii), it remains to be proven that: (a) $\sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta} \times \mathcal{B}} |T^{-1} \widehat{\mathcal{L}}_T(\boldsymbol{\vartheta}, \beta) - \mathcal{L}(\boldsymbol{\vartheta}, \beta)| \xrightarrow{P} 0$, where $\mathcal{L}(\boldsymbol{\vartheta}, \beta) = \mathbb{E}[\ell_t(f_t(\boldsymbol{\vartheta}, \beta), \boldsymbol{\phi})]$ is nonrandom; (b) $\mathcal{L}(\boldsymbol{\vartheta}_0, \beta)$ is independent of $\beta \in \mathcal{B}$. For Part (a), let $\mathcal{L}_T(\boldsymbol{\vartheta}, \beta) = \sum_{t=1}^T \ell_t(f_t(\boldsymbol{\vartheta}, \beta), \boldsymbol{\phi})$. Then,

$$\begin{aligned} & \sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta} \times \mathcal{B}} |T^{-1} \widehat{\mathcal{L}}_T(\boldsymbol{\vartheta}, \beta) - \mathcal{L}(\boldsymbol{\vartheta}, \beta)| \\ & \leq \sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta} \times \mathcal{B}} T^{-1} |\widehat{\mathcal{L}}_T(\boldsymbol{\vartheta}, \beta) - \mathcal{L}_T(\boldsymbol{\vartheta}, \beta)| + \sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta} \times \mathcal{B}} |T^{-1} \mathcal{L}_T(\boldsymbol{\vartheta}, \beta) - \mathcal{L}(\boldsymbol{\vartheta}, \beta)|. \quad (\text{A.1}) \end{aligned}$$

By Assumption 3(i), we note that there exists some $\rho > 1$ such that $\sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta} \times \mathcal{B}} |\ell_t(\widehat{f}_t(\boldsymbol{\vartheta}, \beta), \boldsymbol{\phi}) - \ell_t(f_t(\boldsymbol{\vartheta}, \beta), \boldsymbol{\phi})| \leq C\rho^{-t}$ a.s. for all $t \in \mathbb{Z}^+$. Therefore, it holds a.s. that the first term $\sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta} \times \mathcal{B}} T^{-1} |\widehat{\mathcal{L}}_T(\boldsymbol{\vartheta}, \beta) - \mathcal{L}_T(\boldsymbol{\vartheta}, \beta)| \leq C T^{-1} \sum_{t=1}^T \rho^{-t} \leq C T^{-1} \rightarrow 0$ as $T \rightarrow \infty$. For the second term on the right-hand side of (A.1), $\forall (\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta} \times \mathcal{B}$, we note that $\ell_t(f_t(\boldsymbol{\vartheta}, \beta), \boldsymbol{\phi})$ is a measurable function of $(\mathbf{y}_t, \mathbf{x}_t, f_t(\boldsymbol{\vartheta}, \beta), \boldsymbol{\phi})$ under Assumption 1(iii). By Assumption 2(ii), $f_t(\boldsymbol{\vartheta}, \beta)$ is $\mathcal{F}_{t-1}$ -measurable, which implies that $\{(\mathbf{y}_t, \mathbf{x}_t, f_t(\boldsymbol{\vartheta}, \beta), \boldsymbol{\phi}), t \in \mathbb{Z}\}$ is jointly SE under Assumption 2(i). It then follows from White (2001, Theorem 3.35) that the sequence $\{\ell_t(f_t(\boldsymbol{\vartheta}, \beta), \boldsymbol{\phi}), t \in \mathbb{Z}\}$ is also SE. Since $\ell_t(f_t(\boldsymbol{\vartheta}, \beta), \boldsymbol{\phi})$ is a.s. continuous on $\boldsymbol{\Theta} \times \mathcal{B}$ under Assumption 1(iii) and Assumption 2(ii), and given the compactness of $\boldsymbol{\Theta} \times \mathcal{B}$ (Assumption 1), and the moment condition in Assumption 3(i), the uniform law of large numbers (ULLN) in White (1996, Theorem A.2.2) implies that $\sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta} \times \mathcal{B}} |T^{-1} \mathcal{L}_T(\boldsymbol{\vartheta}, \beta) - \mathcal{L}(\boldsymbol{\vartheta}, \beta)| \xrightarrow{a.s.} 0$, where $\mathcal{L}(\boldsymbol{\vartheta}, \beta)$ is continuous in $(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta} \times \mathcal{B}$. Part (b) follows immediately from the fact that, $\forall \beta \in \mathcal{B}$, we have $f_t(\boldsymbol{\vartheta}_0, \beta) = \omega_0$ if $\alpha_0 = 0$, and hence $\mathcal{L}(\boldsymbol{\vartheta}_0, \beta) =: \mathcal{L}^*(\boldsymbol{\vartheta}_0)$ does not depend on β .

Verification of Assumption 2^{2*}. Let Θ^+ in Andrews (2001, Assumption 2^{2*}) be specified as $\Theta_0^{C,\tilde{\epsilon}}$ for some small $\tilde{\epsilon} > 0$. Then, there exists an $\epsilon > 0$ such that $\Theta \cap \mathcal{S}(\vartheta_0, \epsilon) \subset \Theta_0^{C,\tilde{\epsilon}}$. Assumption 2^{2*} (a) trivially holds (regardless of whether $\alpha_0 = 0$ lies on the boundary or not). Assumption 2^{2*} (b) holds under Assumption 1 and Assumption 2(ii).

We now proceed to verify Assumption 2^{2*} (c). Recall that Θ_0^ϵ is a compact subset of $\Theta \cap \mathcal{S}(\vartheta_0, \epsilon)$. We first check the condition stated in Eq. (9.1) of Andrews (2001), which in our context amounts to showing that

$$\sup_{\vartheta \in \Theta: \|\vartheta - \vartheta_0\| \leq \gamma_T, \beta \in \mathcal{B}} \left| \left(\widehat{\mathcal{L}}_T(\vartheta, \beta) - \widehat{\mathcal{L}}_T(\vartheta_0, \beta) \right) - \left(\mathcal{L}_T(\vartheta, \beta) - \mathcal{L}_T(\vartheta_0, \beta) \right) \right| = o_{\mathbb{P}}(1), \quad (\text{A.2})$$

for all $\gamma_T \rightarrow 0$. Note that the left-hand side of (A.2) is bounded by

$$2 \sum_{t=1}^T \sup_{\vartheta \in \Theta: \|\vartheta - \vartheta_0\| \leq \gamma_T, \beta \in \mathcal{B}} |d_t(\vartheta, \beta)|. \quad (\text{A.3})$$

By Assumption 3(i), the argument below (A.1) shows that $\sum_{t=1}^T \sup_{(\vartheta, \beta) \in \Theta \times \mathcal{B}} |d_t(\vartheta, \beta)| < \infty$. Note that the function $(\vartheta, \beta) \mapsto d_t(\vartheta, \beta)$ is uniformly continuous on $(\vartheta, \beta) \in \Theta \times \mathcal{B}$. This follows from three imposed conditions: first, the uniform continuity of $\hat{f}_t(\cdot)$ and $f_t(\cdot)$ on $\Theta \times \mathcal{B}$ by Assumption 2(ii); second, the continuity of the function $(f, \phi) \mapsto \ell(\mathbf{y}, \mathbf{x}, f, \phi)$ for every $(\mathbf{y}, \mathbf{x}) \in \mathcal{Y} \times \mathcal{X}$ (Assumption 1(iii)); and third, the compactness of $\Theta \times \mathcal{B}$. It then follows that $d_t(\vartheta, \beta)$ is continuous in ϑ uniformly over $\beta \in \mathcal{B}$. Using (A.3) and the same reasoning in Eq. (9.26) of Andrews (2001), one immediately obtains (A.2). As discussed in Andrews (2001, Appendix A), to verify Assumption 2^{2*} in this case, it then suffices to replace $\widehat{\mathcal{L}}_T(\vartheta, \beta)$ with its approximation $\mathcal{L}_T(\vartheta, \beta)$ in Assumption 2^{2*}. Note that

$$\begin{aligned} \sup_{(\vartheta, \beta) \in \Theta_0^\epsilon \times \mathcal{B}} \left\| T^{-1} \frac{\partial^2}{\partial \vartheta \partial \vartheta^\top} \mathcal{L}_T(\vartheta, \beta) - T^{-1} \frac{\partial^2}{\partial \vartheta \partial \vartheta^\top} \mathcal{L}_T(\vartheta_0, \beta) \right\| \\ \leq 2 \sup_{(\vartheta, \beta) \in \Theta_0^\epsilon \times \mathcal{B}} \left\| \left(-T^{-1} \frac{\partial^2}{\partial \vartheta \partial \vartheta^\top} \mathcal{L}_T(\vartheta, \beta) \right) - \mathcal{J}(\vartheta, \beta) \right\|, \end{aligned} \quad (\text{A.4})$$

where $\mathcal{J}(\vartheta, \beta) = -\mathbb{E} \left(\frac{\partial^2}{\partial \vartheta \partial \vartheta^\top} \ell_t(f_t(\vartheta, \beta), \phi) \right)$. Therefore, to verify Assumption 2^{2*} (c), it suffices to show that, for some $\epsilon > 0$,

$$\sup_{(\vartheta, \beta) \in \Theta_0^\epsilon \times \mathcal{B}} \left\| \left(-T^{-1} \frac{\partial^2}{\partial \vartheta \partial \vartheta^\top} \mathcal{L}_T(\vartheta, \beta) \right) - \mathcal{J}(\vartheta, \beta) \right\| = o_{a.s.}(1). \quad (\text{A.5})$$

For $(\vartheta, \beta) \in \Theta_0^\epsilon \times \mathcal{B}$, $\frac{\partial^2}{\partial \vartheta \partial \vartheta^\top} \ell_t(f_t(\vartheta, \beta), \phi)$ is a measurable function of

$(\mathbf{y}_t, \mathbf{x}_t, f_t(\boldsymbol{\vartheta}, \beta), \frac{\partial}{\partial \boldsymbol{\vartheta}} f_t(\boldsymbol{\vartheta}, \beta), \frac{\partial^2}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}^\top} f_t(\boldsymbol{\vartheta}, \beta))$ and is also continuous in $(\boldsymbol{\vartheta}, \beta)$ under [Assumption 1](#) and [Assumption 2\(ii\)](#). Then, (A.5) follows directly from [Assumption 3\(iii\)](#), together with the ULLN in [White \(1996, Theorem A.2.2\)](#), applied to the SE process $\left\{ \frac{\partial^2}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}^\top} \ell_t(f_t(\boldsymbol{\vartheta}, \beta), \boldsymbol{\phi}), t \in \mathbb{Z} \right\}$ whose SE property is ensured by [Assumption 2\(ii\)](#).

Verification of Assumption 3*. Since $f_t(\boldsymbol{\vartheta}_0, \beta) = \hat{f}_t(\boldsymbol{\vartheta}_0, \beta) = \omega_0$ for all $t \geq 1$ and $\beta \in \mathcal{B}$, and $\mathbb{E}(\log^+ |\nabla_t^f(\omega_0, \boldsymbol{\phi}_0)|) < \infty$, it follows from [Straumann and Mikosch \(2006, Lemma 2.1\)](#) and [Assumption 2\(ii\)](#) that $T^{-1/2} \partial \hat{\mathcal{L}}_T(\boldsymbol{\vartheta}, \beta) / \partial \boldsymbol{\vartheta} \big|_{\boldsymbol{\vartheta}=\boldsymbol{\vartheta}_0} = T^{-1/2} \partial \mathcal{L}_T(\boldsymbol{\vartheta}, \beta) / \partial \boldsymbol{\vartheta} \big|_{\boldsymbol{\vartheta}=\boldsymbol{\vartheta}_0} + o_{a.s.}(1)$. We therefore establish the weak convergence of $T^{-1/2} \partial \mathcal{L}_T(\boldsymbol{\vartheta}, \cdot) / \partial \boldsymbol{\vartheta} \big|_{\boldsymbol{\vartheta}=\boldsymbol{\vartheta}_0}$ as a process indexed by $\beta \in \mathcal{B}$. Recall from [Theorem 1](#) $h_{0,t}(\beta) = \sum_{j=0}^{\infty} \beta^j s_{t-j}(f, \boldsymbol{\phi}) \big|_{(f, \boldsymbol{\phi})=(\omega_0, \boldsymbol{\phi}_0)}$ and note that $\partial f_t(\boldsymbol{\vartheta}, \beta) / \partial \boldsymbol{\vartheta} \big|_{\boldsymbol{\vartheta}=\boldsymbol{\vartheta}_0} = (h_{0,t-1}(\beta), 1, \mathbf{0}^\top)^\top$. Let $\nabla_{0,t}^f := \nabla_t^f(f, \boldsymbol{\phi}) \big|_{(f, \boldsymbol{\phi})=(\omega_0, \boldsymbol{\phi}_0)}$ and $\nabla_{0,t}^\phi := \nabla_t^\phi(f, \boldsymbol{\phi}) \big|_{(f, \boldsymbol{\phi})=(\omega_0, \boldsymbol{\phi}_0)}$, we can write

$$\begin{aligned} & \frac{1}{\sqrt{T}} \frac{\partial}{\partial \boldsymbol{\vartheta}} \mathcal{L}_T(\boldsymbol{\vartheta}, \beta) \bigg|_{\boldsymbol{\vartheta}=\boldsymbol{\vartheta}_0} \\ &= \frac{1}{\sqrt{T}} \sum_{t=1}^T \left(\nabla_t^f(f, \boldsymbol{\phi}) \bigg|_{(f, \boldsymbol{\phi})=(\omega_0, \boldsymbol{\phi}_0)} \frac{\partial f_t(\boldsymbol{\vartheta}, \beta)}{\partial \boldsymbol{\vartheta}} \bigg|_{\boldsymbol{\vartheta}=\boldsymbol{\vartheta}_0} + \frac{\partial \log p(\mathbf{y}_t | \mathbf{x}_t, f, \boldsymbol{\phi})}{\partial \boldsymbol{\vartheta}} \bigg|_{(f, \boldsymbol{\phi})=(\omega_0, \boldsymbol{\phi}_0)} \right) \\ &= \frac{1}{\sqrt{T}} \sum_{t=1}^T \begin{pmatrix} \nabla_{0,t}^f h_{0,t-1}(\beta) \\ \nabla_{0,t}^f \\ \nabla_{0,t}^\phi \end{pmatrix} =: \frac{1}{\sqrt{T}} \sum_{t=1}^T \mathbf{g}_{0,t}(\beta). \end{aligned} \tag{A.6}$$

We employ Theorem 10.2 of [Pollard \(1990\)](#) to derive the weak convergence of $T^{-1/2} \sum_{t=1}^T \mathbf{g}_t(\cdot)$. This requires verifying three conditions: (a) $\mathcal{B}$ is totally bounded; (b) the finite dimensional distributions of $T^{-1/2} \sum_{t=1}^T \mathbf{g}_t(\cdot)$ converge to a limiting distribution; (c) $\{T^{-1/2} \sum_{t=1}^T \mathbf{g}_t(\cdot), T \geq 1\}$ is stochastically equicontinuous.

Condition (a) holds because $\mathcal{B} \subset (-1, 1)$. Now we prove Condition (b). Under [Assumption 3\(iv\)](#), we see that $\sum_{t=1}^T \mathbf{g}_{0,t}(\beta)$ is a partial sum of SE martingale difference (m.d.) random variables. We can apply the Cramér-Wold device together with a central limit theorem (CLT) for SE m.d. sequences, such as Theorem 24.3 of [Davidson \(1994\)](#) (see also Theorems 13.12 and 23.16, and the discussion in Chapter 24.3, p. 385 of the same book), or Theorem 18.3 of [Billingsley \(1999\)](#), provided that $\mathbb{E}(\sup_{\beta \in \mathcal{B}} \|\mathbf{g}_{0,t}(\beta)\|^2) < \infty$. Since $\mathbf{g}_{0,t}(\beta)$ has fixed dimension, it suffices to consider its components individually. Let $\bar{\beta}$ be a number strictly between $\sup\{|\beta|, \beta \in \mathcal{B}\}$ and

1. For the first component of $\mathbf{g}_{0,t}(\beta)$, by [Assumption 3\(iv\)](#) and the Cauchy-Schwarz inequality,

$$\begin{aligned}\mathbb{E}\left([\nabla_{0,t}^f]^2 \sup_{\beta \in \mathcal{B}} [h_{0,t-1}(\beta)]^2\right) &= \Sigma_{ff} \mathbb{E}\left(\sup_{\beta \in \mathcal{B}} \left(\sum_{j=0}^{\infty} \frac{\beta^j}{\bar{\beta}^j} \bar{\beta}^j s_{t-1-j}(\omega_0, \phi_0)\right)^2\right) \\ &\leq \Sigma_{ff} \left(\sup_{\beta \in \mathcal{B}} \sum_{j=0}^{\infty} \left(\frac{\beta}{\bar{\beta}}\right)^{2j}\right) \sum_{j=0}^{\infty} \bar{\beta}^{2j} \mathbb{E}(s_{t-1-j}^2(\omega_0, \phi_0)) < \infty,\end{aligned}\tag{A.7}$$

by the strict stationarity of $\{s_t(\omega_0, \phi_0), t \in \mathbb{Z}\}$. For the remaining components of $\mathbf{g}_{0,t}(\beta)$ that do not depend on β , the existence of second-order moments follows directly from [Assumption 3\(iv\)](#). Applying a CLT for SE m.d. sequences as described above, we obtain that each of the finite-dimensional distributions of $T^{-1/2} \sum_{t=1}^T \mathbf{g}_{0,t}(\cdot)$ converges in distribution to a multivariate normal distribution with covariance determined by the covariance function in [\(3.4\)](#). Consider Condition (c) where we need to establish the stochastic equicontinuity of $\{T^{-1/2} \sum_{t=1}^T \mathbf{g}_t(\cdot), T \geq 1\}$. As above, applying the Cauchy-Schwarz inequality, we note that for any $\delta > 0$,

$$\begin{aligned}&\mathbb{E}\left\{\sup_{|\beta_1 - \beta_2| < \delta} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T \nabla_{0,t}^f h_{0,t-1}(\beta_1) - \frac{1}{\sqrt{T}} \sum_{t=1}^T \nabla_{0,t}^f h_{0,t-1}(\beta_2)\right)^2\right\} \\ &= \mathbb{E}\left\{\sup_{|\beta_1 - \beta_2| < \delta} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T \nabla_{0,t}^f \left(\sum_{j=0}^{\infty} (\beta_1^j - \beta_2^j) s_{t-1-j}(\omega_0, \phi_0)\right)\right)^2\right\} \\ &= \mathbb{E}\left\{\sup_{|\beta_1 - \beta_2| < \delta} \left(\sum_{j=0}^{\infty} \frac{\beta_1^j - \beta_2^j}{\bar{\beta}^j} \bar{\beta}^j \frac{1}{\sqrt{T}} \sum_{t=1}^T \nabla_{0,t}^f s_{t-1-j}(\omega_0, \phi_0)\right)^2\right\} \\ &\leq \left(\sup_{|\beta_1 - \beta_2| < \delta} \sum_{j=0}^{\infty} \frac{(\beta_1^j - \beta_2^j)^2}{\bar{\beta}^{2j}}\right) \left\{\sum_{j=0}^{\infty} \bar{\beta}^{2j} \mathbb{E}\left(\frac{1}{\sqrt{T}} \sum_{t=1}^T \nabla_{0,t}^f s_{t-1-j}(\omega_0, \phi_0)\right)^2\right\} \\ &= \frac{\Sigma_{ff} \mathbb{E}(s_1(\omega_0, \phi_0))^2}{\beta_U^2} \sup_{|\beta_1 - \beta_2| < \delta} \sum_{j=0}^{\infty} \frac{(\beta_1^j - \beta_2^j)^2}{\bar{\beta}^{2j}},\end{aligned}$$

where the final step follows from

$$\mathbb{E}\left(\frac{1}{\sqrt{T}} \sum_{t=1}^T \nabla_{0,t}^f s_{t-1-j}(\omega_0, \phi_0)\right)^2 = \frac{1}{T} \sum_{t=1}^T \mathbb{E}[(\nabla_{0,t}^f)^2] \mathbb{E}[s_{t-1-j}^2(\omega_0, \phi_0)] = \Sigma_{ff} \mathbb{E}[s_1^2(\omega_0, \phi_0)].$$

The stochastic equicontinuity of $\{T^{-1/2} \sum_{t=1}^T \mathbf{g}_t(\cdot), T \geq 1\}$ can then be established using the same steps as in [Andrews and Ploberger \(1996, p. 1340, below \(A.14\)\)](#). Combining these three conditions yields [\(3.3\)](#) immediately.

Moreover, note that $\mathcal{J}(\boldsymbol{\vartheta}_0, \beta) = -\mathbb{E}\left(\frac{\partial^2}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}^\top} \ell_t(f_t(\boldsymbol{\vartheta}_0, \beta), \phi_0)\right)$ is symmetric and nonrandom.

It remains to be shown that

$$0 < \inf_{\beta \in \mathcal{B}} \lambda_{\min}(\mathcal{J}(\boldsymbol{\vartheta}_0, \beta)) \leq \sup_{\beta \in \mathcal{B}} \lambda_{\max}(\mathcal{J}(\boldsymbol{\vartheta}_0, \beta)) < \infty, \quad (\text{A.8})$$

where $\lambda_{\min}(\cdot)$ and $\lambda_{\max}(\cdot)$ denote the smallest and largest eigenvalues, respectively. Note that

$$\frac{\partial^2}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}^\top} \ell_t(f_t(\boldsymbol{\vartheta}, \beta), \boldsymbol{\phi}) = \sum_{k=1}^3 \mathbf{L}_{k,t}(\boldsymbol{\vartheta}, \beta), \text{ where}$$

$$\begin{aligned} \mathbf{L}_{1,t}(\boldsymbol{\vartheta}, \beta) &= \frac{\partial \nabla_t^f(f, \boldsymbol{\phi})}{\partial f} \Big|_{f=f_t(\boldsymbol{\vartheta}, \beta)} \frac{\partial f_t(\boldsymbol{\vartheta}, \beta)}{\partial \boldsymbol{\vartheta}} \frac{\partial f_t(\boldsymbol{\vartheta}, \beta)}{\partial \boldsymbol{\vartheta}^\top} + \frac{\partial^2 \ell_t(f, \boldsymbol{\phi})}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}^\top} \Big|_{f=f_t(\boldsymbol{\vartheta}, \beta)}, \\ \mathbf{L}_{2,t}(\boldsymbol{\vartheta}, \beta) &= \frac{\partial \nabla_t^f(f, \boldsymbol{\phi})}{\partial \boldsymbol{\vartheta}} \Big|_{f=f_t(\boldsymbol{\vartheta}, \beta)} \frac{\partial f_t(\boldsymbol{\vartheta}, \beta)}{\partial \boldsymbol{\vartheta}^\top} + \left(\frac{\partial \nabla_t^f(f, \boldsymbol{\phi})}{\partial \boldsymbol{\vartheta}} \Big|_{f=f_t(\boldsymbol{\vartheta}, \beta)} \frac{\partial f_t(\boldsymbol{\vartheta}, \beta)}{\partial \boldsymbol{\vartheta}^\top} \right)^\top, \\ \mathbf{L}_{3,t}(\boldsymbol{\vartheta}, \beta) &= \nabla_t^f(f, \boldsymbol{\phi}) \Big|_{f=f_t(\boldsymbol{\vartheta}, \beta)} \frac{\partial^2 f_t(\boldsymbol{\vartheta}, \beta)}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}^\top}. \end{aligned}$$

For brevity, let $\nabla_{0,t}^{ff} := \frac{\partial \nabla_t^f(f, \boldsymbol{\phi})}{\partial f} \Big|_{(f, \boldsymbol{\phi})=(\omega_0, \boldsymbol{\phi}_0)}$, $\nabla_{0,t}^{\phi\phi} := \frac{\partial}{\partial \boldsymbol{\phi}} \nabla_t^\phi(f, \boldsymbol{\phi})^\top \Big|_{(f, \boldsymbol{\phi})=(\omega_0, \boldsymbol{\phi}_0)}$, and $\nabla_{0,t}^{\phi f} = \nabla_t^{\phi f}(\omega_0, \boldsymbol{\phi}_0)$. We then obtain

$$\mathbf{L}_{1,t}(\boldsymbol{\vartheta}_0, \beta) + \mathbf{L}_{2,t}(\boldsymbol{\vartheta}_0, \beta) = \begin{pmatrix} \nabla_{0,t}^{ff} h_{0,t-1}^2(\beta) & \nabla_{0,t}^{ff} h_{0,t-1}(\beta) & \nabla_{0,t}^{\phi f \top} h_{0,t-1}(\beta) \\ \nabla_{0,t}^{ff} h_{0,t-1}(\beta) & \nabla_{0,t}^{ff} & \nabla_{0,t}^{\phi f \top} \\ \nabla_{0,t}^{\phi f \top} h_{0,t-1}(\beta) & \nabla_{0,t}^{\phi f} & \nabla_{0,t}^{\phi\phi} \end{pmatrix},$$

and $\mathbf{L}_{3,t}(\boldsymbol{\vartheta}_0, \beta) = \nabla_t^f(\omega_0, \boldsymbol{\phi}_0) \frac{\partial^2 f_t(\boldsymbol{\vartheta}, \beta)}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}^\top} \Big|_{\boldsymbol{\vartheta}=\boldsymbol{\vartheta}_0}$. Note that $\frac{\partial^2 f_t(\boldsymbol{\vartheta}, \beta)}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}^\top} \Big|_{\boldsymbol{\vartheta}=\boldsymbol{\vartheta}_0}$ is $\mathcal{F}_{t-1}$ -measurable (Assumption 2(ii)). Hence, $\mathbb{E}(\mathbf{L}_{3,t}(\boldsymbol{\vartheta}_0, \beta)) = \mathbf{0}$ for all $\beta \in \mathcal{B}$ by Assumption 3(iv). If the condition $\mathbb{E}(\nabla_{0,t}^{\phi f} \mid \mathcal{F}_{t-1}) = \boldsymbol{\Omega}_{\phi f} = \mathbf{0}$ in Assumption 3(iv) holds, then

$$\mathbb{E}(\mathbf{L}_{1,t}(\boldsymbol{\vartheta}_0, \beta) + \mathbf{L}_{2,t}(\boldsymbol{\vartheta}_0, \beta)) = \begin{pmatrix} \mathbb{E}(\nabla_{0,t}^{ff}) \mathbb{E}(h_{0,t-1}^2(\beta)) & \mathbb{E}(\nabla_{0,t}^{ff}) \mathbb{E}(h_{0,t-1}(\beta)) & \mathbf{0}^\top \\ \mathbb{E}(\nabla_{0,t}^{ff}) \mathbb{E}(h_{0,t-1}(\beta)) & \mathbb{E}(\nabla_{0,t}^{ff}) & \mathbf{0}^\top \\ \mathbf{0} & \mathbf{0} & \mathbb{E}(\nabla_{0,t}^{\phi\phi}) \end{pmatrix},$$

and thus $\mathcal{J}(\boldsymbol{\vartheta}_0, \beta) = \text{diag}(\mathcal{J}_{(\alpha, \omega)}(\beta), \mathcal{J}_{\phi\phi})$, where $\mathcal{J}_{(\alpha, \omega)}(\beta)$ is defined in Eq. (3.6) and $\mathcal{J}_{\phi\phi} := -\mathbb{E}(\nabla_{0,t}^{\phi\phi})$. Note that $\sup_{\beta \in \mathcal{B}} \mathbb{E}([h_{0,t-1}(\beta)]^2) < \infty$ as seen in (A.7). It then follows from Assumption 3(v) that (A.8) holds. Alternatively, if $\boldsymbol{\Omega}_{\phi f}$ is possibly nonzero but the condition $\mathbb{E}(h_{0,t}(\beta)) = 0$ in Assumption 3(iv) holds, then one can write $\mathcal{J}(\boldsymbol{\vartheta}_0, \beta) = \text{diag}(\mathcal{J}_\alpha(\beta), \mathcal{J}_{(\omega, \phi)})$, where $\mathcal{J}_\alpha(\beta) = -\mathbb{E}(\nabla_{0,t}^{ff}) \mathbb{E}(h_{0,t-1}^2(\beta))$ is the (1,1)-entry of $\mathcal{J}_{(\alpha, \omega)}(\beta)$ (also defined in the theorem), and $\mathcal{J}_{(\omega, \phi)} := -\mathbb{E} \begin{pmatrix} \nabla_{0,t}^{ff} & \nabla_{0,t}^{\phi f \top} \\ \nabla_{0,t}^{\phi f} & \nabla_{0,t}^{\phi\phi} \end{pmatrix}$. Since $\mathcal{J}_{(\omega, \phi)}$ is positive definite by Assumption 3(v), thereby completing the verification of Assumption 3*.

Verification of Assumptions 5* and 7–10. Consider Assumption 5*. We focus on the case $\phi_0 \in \text{int}(\Theta_\phi)$. Let $\mathbf{A} := \Lambda_\alpha \times \Lambda_\omega \times \mathbf{A}_\phi$, where $\Lambda_\omega = \mathbb{R}$ and $\mathbf{A}_\phi = \mathbb{R}^m$. Recall that $\Theta_\alpha = [\alpha_L, \alpha_U] \subset \mathbb{R}$ for some $\alpha_U > 0$. We set $\Lambda_\alpha = \mathbb{R}^+$ if $\alpha_L = 0$, and $\Lambda_\alpha = \mathbb{R}$ if $\alpha_L < 0$. With this choice of $\mathbf{A}$, Assumption 5*(a) is satisfied. Assumption 5*(b) is immediate. Note that if some or all elements of ϕ_0 lie on the boundary of Θ_ϕ , the above arguments can be adapted by restricting the cone $\mathbf{A}_\phi$ to the corresponding positive or negative half-space for those elements. To verify Assumption 7(a), we partition the parameter vector $\boldsymbol{\vartheta} = (\alpha, \omega, \phi^\top)^\top$ depending on whether $\boldsymbol{\Omega}_{\phi_f} = \mathbf{0}$ or not. If $\boldsymbol{\Omega}_{\phi_f} = \mathbf{0}$, we take $\theta_* = (\alpha, \omega)^\top$ and $\psi = \boldsymbol{\vartheta}$ in the notation of Andrews (2001). If $\boldsymbol{\Omega}_{\phi_f}$ is possibly nonzero, but $\mathbb{E}(h_{0,t}(\beta)) = 0$ holds, we instead set $\theta_* = \alpha$ and $\psi = (\omega, \phi^\top)^\top$. In either case, Assumption 7(a) is satisfied due to the block-diagonal structure of $\mathcal{J}(\boldsymbol{\vartheta}_0, \beta)$ established above. Furthermore, Assumption 7(b) holds because $\mathbf{A}$ is a product set by construction. Assumption 8 also holds by the construction of $\mathbf{A}_\phi$. Assumption 9(a) holds by our definition of Θ_0 , and 9(b)–(d) are straightforward. Assumption 10(a) holds because the initial condition is set as $\hat{f}_1(\boldsymbol{\vartheta}_0, \beta) = \omega_0$; it would not hold otherwise. Assumption 10(b) is satisfied since only the first component in (A.6) and $\mathcal{J}_\alpha(\beta)$ depend on β , while $\mathcal{J}_{(\omega, \phi)}$ and $\mathcal{J}_{\phi\phi}$ defined above are independent of β .

Eq. (3.5) follows from Theorem 4(b)–(c) of Andrews (2001), in combination with Theorem 2(a) of Andrews (2001) and Theorem 5 of Andrews (1999). Note that for $\Theta_\alpha = [0, \alpha_U]$, one can also refer to Eq. (3.10) of Andrews (2001). $\square$

Online Appendix to:

TESTING FOR THE ABSENCE OF SCORE-DRIVEN PARAMETER DYNAMICS

Yicong Lin¹ and Andre Lucas¹

¹ *Vrije Universiteit Amsterdam and Tinbergen Institute*

Contents

A Additional proofs and results	Appendix p. 2
B Simulated critical values	Appendix p. 11
C Additional details on the four examples	Appendix p. 13
D Parameter estimates for the empirical example	Appendix p. 22

A Additional proofs and results

Proof of Corollary 1. Note that $\mathbb{E}(h_{0,t-1}(\beta)) = 0$ for all $\beta \in \mathcal{B}$. Thus, the results in Part (ii) of Theorem 1 apply directly. It follows that for every $\beta \in \mathcal{B}$, $Z_\alpha(\beta) := \mathcal{J}_\alpha^{-1}(\beta) G_\alpha(\beta) \sim \mathcal{N}(0, \kappa_Z(1 - \beta^2))$, where $\kappa_Z := \kappa_{\mathcal{J}}^{-1} \kappa_G$, $\kappa_{\mathcal{J}} := \Omega_{ff} \mathbb{E}(s^2(\mathbf{y}_t, \mathbf{x}_t, \omega_0, \phi_0))$, and $\kappa_G := \Omega_{ff}^{-1} \Sigma_{ff}$. The process Z_α is therefore a Gaussian process with the covariance function given by

$$\text{Cov}(Z_\alpha(\beta_1), Z_\alpha(\beta_2)) = \kappa_Z \frac{(1 - \beta_1^2)(1 - \beta_2^2)}{1 - \beta_1 \beta_2}. \quad (\text{A.1})$$

One can verify that $\kappa_Z^{1/2}(1 - \beta^2) \sum_{j=0}^{\infty} \beta^j X_j$ defines a Gaussian process whose covariance function coincides with that of Z_α in (A.1). Hence, $Z_\alpha(\beta) \stackrel{d}{=} \kappa_Z^{1/2}(1 - \beta^2) \sum_{j=0}^{\infty} \beta^j X_j$, where $\stackrel{d}{=}$ denotes equality in distribution. Note that $\mathcal{J}_\alpha(\beta) = \kappa_{\mathcal{J}}(1 - \beta^2)^{-1}$. By Theorem 1, we obtain

$$QLR_T \xrightarrow{d} \begin{cases} \kappa_G \sup_{\beta \in \mathcal{B}} \left(\max \left\{ (1 - \beta^2)^{1/2} \sum_{j=0}^{\infty} \beta^j X_j, 0 \right\} \right)^2, & \alpha_L = 0, \\ \kappa_G \sup_{\beta \in \mathcal{B}} \left((1 - \beta^2)^{1/2} \sum_{j=0}^{\infty} \beta^j X_j \right)^2, & \alpha_L < 0. \end{cases} \quad (\text{A.2})$$

Then (3.7) immediately follows from (ii). $\square$

Proof of Corollary 2. This directly follows from (A.2), noting that $\kappa_G = 1$. $\square$

The following proofs mainly draw on the results of Bougerol (1993) and Straumann and Mikosch (2006). Consider a complete separable metric space (E, d_E) , in line with Bougerol (1993, Section 3), and define the Lipschitz coefficient ρ associated with a random map $\varphi : E \rightarrow E$ as:

$$\rho(\varphi) = \sup_{x, y \in E, x \neq y} \left\{ \frac{d_E(\varphi(x), \varphi(y))}{d_E(x, y)} \right\}. \quad (\text{A.3})$$

Let $\mathcal{C}^0(\boldsymbol{\Theta} \times \mathcal{B}, \mathcal{F})$ denote the space of continuous $\mathcal{F}$ -valued functions on $\boldsymbol{\Theta} \times \mathcal{B}$, equipped with the supremum norm $\|\cdot\|_{\boldsymbol{\Theta} \times \mathcal{B}}$, defined by $\|f\|_{\boldsymbol{\Theta} \times \mathcal{B}} = \sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta} \times \mathcal{B}} |f(\boldsymbol{\vartheta}, \beta)|$ for $f \in \mathcal{C}^0(\boldsymbol{\Theta} \times \mathcal{B}, \mathcal{F})$. Then, $(\mathcal{C}^0(\boldsymbol{\Theta} \times \mathcal{B}, \mathcal{F}), \|\cdot\|_{\boldsymbol{\Theta} \times \mathcal{B}})$ is a complete and separable (and thus Polish) space. Moreover, in the proofs below, we repeatedly use the following inequalities without further mention (see, e.g., Lin and Lucas, 2025, (B.3)–(B.4)): for any matrices $\mathbf{X}_t$, $t = 1, \dots, K$, where $K \in \mathbb{Z}^+$, with

compatible dimensions, we have

$$\log^+ \left\| \prod_{t=1}^K \mathbf{X}_t \right\| \leq \log^+ \left(\prod_{t=1}^K \|\mathbf{X}_t\| \right) \leq \sum_{t=1}^K \log^+ \|\mathbf{X}_t\|, \quad (\text{A.4})$$

$$\log^+ \left\| \sum_{t=1}^K \mathbf{X}_t \right\| \leq \log^+ \left(\sum_{t=1}^K \|\mathbf{X}_t\| \right) \leq \log(K) + \sum_{t=1}^K \log^+ \|\mathbf{X}_t\|. \quad (\text{A.5})$$

We use C to denote a generic positive constant that may vary from line to line.

Proof of Proposition 1. We shall employ Theorem 2.10 of [Straumann and Mikosch \(2006\)](#) for *perturbed*, nonstationary SREs. Before proceeding, recall that $\mathbf{f}_t^{(0)}(\boldsymbol{\vartheta}, \beta) = f_t(\boldsymbol{\vartheta}, \beta)$ and $\hat{\mathbf{f}}_t^{(0)}(\boldsymbol{\vartheta}, \beta) := \hat{f}_t(\boldsymbol{\vartheta}, \beta)$. Note that for $k \geq 1$, one can generally write (provided the partial derivatives with respect to $\boldsymbol{\vartheta}$ are well-defined) $\hat{\mathbf{f}}_{t+1}^{(k)} = \hat{\boldsymbol{\varphi}}_{(k),t}(\hat{\mathbf{f}}_t^{(k)})$, where $\hat{\boldsymbol{\varphi}}_{(k),t}$ is given by

$$\begin{aligned} & [\hat{\boldsymbol{\varphi}}_{(k),t}(\mathbf{f}^{(k)})](\boldsymbol{\vartheta}, \beta) \\ &:= \hat{\boldsymbol{\psi}}_{(k),t}(\mathbf{f}^{(k)}(\boldsymbol{\vartheta}, \beta), \boldsymbol{\vartheta}, \beta) \\ &:= Q_t(\hat{\mathbf{f}}_t^{(0)}(\boldsymbol{\vartheta}, \beta), \boldsymbol{\vartheta}, \beta) \mathbf{f}^{(k)}(\boldsymbol{\vartheta}, \beta) + \mathbf{q}_{(k),t}(\hat{\mathbf{f}}_t^{(k-1)}(\boldsymbol{\vartheta}, \beta), \dots, \hat{\mathbf{f}}_t^{(0)}(\boldsymbol{\vartheta}, \beta), \boldsymbol{\vartheta}, \beta), \end{aligned} \quad (\text{A.6})$$

where Q_t is a scalar function that depends on the initialized filter $\hat{f}_t(\boldsymbol{\vartheta}, \beta)$, and $\mathbf{q}_{(k),t}$ for $k \geq 1$ is a vector-valued function that depends only on its derivatives up to order $k - 1$. The (perturbed) derivative process $\{\hat{\mathbf{f}}_t^{(k)}(\cdot), t \in \mathbb{Z}^+\}$ depends on the initialized sequence $\{\hat{\mathbf{f}}_t^{(0)}(\cdot), \hat{\mathbf{f}}_t^{(1)}(\cdot), \dots, \hat{\mathbf{f}}_t^{(k-1)}(\cdot), t \in \mathbb{Z}^+\}$ and is therefore nonstationary. To employ Theorem 2.10 of [Straumann and Mikosch \(2006\)](#), define for $k \geq 1$ an intermediate *unperturbed* sequence $\{\hat{\mathbf{d}}_t^{(k)}(\cdot), t \in \mathbb{Z}^+\}$, initialized at some initial function $\hat{\mathbf{f}}_1^{(k)}(\cdot)$ and depending solely on the limit sequence $\{\mathbf{f}_t^{(0)}(\cdot), \mathbf{f}_t^{(1)}(\cdot), \dots, \mathbf{f}_t^{(k-1)}(\cdot), t \in \mathbb{Z}\}$, which is associated with the random maps $\boldsymbol{\varphi}_{(k),t}$ defined by

$$\begin{aligned} & [\boldsymbol{\varphi}_{(k),t}(\mathbf{d}^{(k)})](\boldsymbol{\vartheta}, \beta) \\ &:= \boldsymbol{\psi}_{(k),t}(\mathbf{d}^{(k)}(\boldsymbol{\vartheta}, \beta), \boldsymbol{\vartheta}, \beta) \\ &:= Q_t(\mathbf{f}_t^{(0)}(\boldsymbol{\vartheta}, \beta), \boldsymbol{\vartheta}, \beta) \mathbf{d}^{(k)}(\boldsymbol{\vartheta}, \beta) + \mathbf{q}_{(k),t}(\mathbf{f}_t^{(k-1)}(\boldsymbol{\vartheta}, \beta), \dots, \mathbf{f}_t^{(0)}(\boldsymbol{\vartheta}, \beta), \boldsymbol{\vartheta}, \beta). \end{aligned} \quad (\text{A.7})$$

For the specific cases of $k = 1, 2$ here, it follows by mathematical induction and [Assumption 1\(iii\)](#) that, for all $t \in \mathbb{Z}^+$ and $(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta}_0^\epsilon \times \mathcal{B}$, the initialized sequence $\hat{f}_t(\boldsymbol{\vartheta}, \beta)$, starting from the initial value $\hat{f}_1(\boldsymbol{\vartheta}, \beta) = \omega$, admits *continuous* l/r partial derivatives of order two with respect to $\boldsymbol{\vartheta}$, and these partial derivatives are continuous in $(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta}_0^\epsilon \times \mathcal{B}$. The specific

construction in (A.6) reduces to

$$\hat{\mathbf{f}}_{t+1}^{(1)}(\boldsymbol{\vartheta}, \beta) = Q_t(\hat{f}_t(\boldsymbol{\vartheta}, \beta), \boldsymbol{\vartheta}, \beta) \hat{\mathbf{f}}_t^{(1)}(\boldsymbol{\vartheta}, \beta) + \mathbf{q}_{(1),t}(\hat{f}_t(\boldsymbol{\vartheta}, \beta), \boldsymbol{\vartheta}, \beta), \quad (\text{A.8})$$

$$\hat{\mathbf{f}}_{t+1}^{(2)}(\boldsymbol{\vartheta}, \beta) = Q_t(\hat{f}_t(\boldsymbol{\vartheta}, \beta), \boldsymbol{\vartheta}, \beta) \hat{\mathbf{f}}_t^{(2)}(\boldsymbol{\vartheta}, \beta) + \mathbf{q}_{(2),t}(\hat{\mathbf{f}}_t^{(1)}(\boldsymbol{\vartheta}, \beta), \hat{f}_t(\boldsymbol{\vartheta}, \beta), \boldsymbol{\vartheta}, \beta), \quad (\text{A.9})$$

where $Q_t(f, \boldsymbol{\vartheta}, \beta) = \beta + \alpha \frac{\partial s_t(f, \phi)}{\partial f}$, $\mathbf{q}_{(1),t}(f, \boldsymbol{\vartheta}, \beta) = \left(s_t(f, \phi), 1 - \beta, \alpha \frac{\partial s_t(f, \phi)}{\partial \phi^\top} \right)^\top$, and

$$\begin{aligned} \mathbf{q}_{(2),t}(\mathbf{f}^{(1)}, f, \boldsymbol{\vartheta}, \beta) = & \text{vec} \left(\mathbf{f}^{(1)} \frac{\partial \alpha}{\partial \boldsymbol{\vartheta}^\top} \frac{\partial s_t(f, \phi)}{\partial f} + \alpha \mathbf{f}^{(1)} \frac{\partial^2 s_t(f, \phi)}{\partial \boldsymbol{\vartheta}^\top \partial f} + \alpha \frac{\partial^2 s_t(f, \phi)}{\partial f^2} \mathbf{f}^{(1)} \mathbf{f}^{(1)\top} \right. \\ & + \frac{\partial s_t(f, \phi)}{\partial f} \frac{\partial \alpha}{\partial \boldsymbol{\vartheta}} \mathbf{f}^{(1)\top} + \alpha \frac{\partial^2 s_t(f, \phi)}{\partial f \partial \boldsymbol{\vartheta}} \mathbf{f}^{(1)\top} + \frac{\partial^2 s_t(f, \phi)}{\partial \boldsymbol{\vartheta}^\top \partial \boldsymbol{\vartheta}} \\ & \left. + \frac{\partial \alpha}{\partial \boldsymbol{\vartheta}} \frac{\partial s_t(f, \phi)}{\partial \boldsymbol{\vartheta}^\top} + \frac{\partial s_t(f, \phi)}{\partial \boldsymbol{\vartheta}} \frac{\partial \alpha}{\partial \boldsymbol{\vartheta}^\top} \right). \end{aligned}$$

For the unperturbed counterparts (see (A.7)), the expressions for $k = 1, 2$ are identical, except for plugging in $\mathbf{f}_t^{(1)}(\boldsymbol{\vartheta}, \beta)$ and $f_t(\boldsymbol{\vartheta}, \beta)$ instead of $\hat{\mathbf{f}}_t^{(1)}(\boldsymbol{\vartheta}, \beta)$ and $\hat{f}_t(\boldsymbol{\vartheta}, \beta)$, respectively, in Q_t and $\mathbf{q}_{(k),t}$ above. Note that the random maps $\hat{\boldsymbol{\varphi}}_{(k),t}$ and $\boldsymbol{\varphi}_{(k),t}$ for $k \geq 1$ are defined on the complete and separable metric space $(\mathcal{C}^0(\boldsymbol{\Theta}_0^\epsilon \times \mathcal{B}, \mathbb{R}^{(m+2)^k}), \|\cdot\|_{\boldsymbol{\Theta}_0^\epsilon \times \mathcal{B}})$, where $\|\mathbf{f}\|_{\boldsymbol{\Theta}_0^\epsilon \times \mathcal{B}} = \sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta}_0^\epsilon \times \mathcal{B}} \|\mathbf{f}(\boldsymbol{\vartheta}, \beta)\|$. Similar to Blasques et al. (2022, Proposition 3.4) and Lin and Lucas (2025, Proposition 4), the following high-level conditions (abbreviated as $\widehat{\text{HL}}$) are sufficient for applying Straumann and Mikosch (2006, Theorem 2.10): For $k \in \mathbb{Z}^+$, the sequence $\{\boldsymbol{\varphi}_{(k),t}, t \in \mathbb{Z}\}$ is SE. Moreover,

$$\widehat{\text{HL1}} \quad \mathbb{E} \left(\log^+ \|\boldsymbol{\varphi}_{(k),1}(\hat{\mathbf{f}}_1^{(k)})\|_{\boldsymbol{\Theta}_0^\epsilon \times \mathcal{B}} \right) < \infty, \text{ where } \hat{\mathbf{f}}_1^{(1)}(\boldsymbol{\vartheta}, \beta) = (0, 1, 0, \dots, 0)^\top \in \mathbb{R}^{m+2}, \text{ and } \hat{\mathbf{f}}_1^{(k)}(\boldsymbol{\vartheta}, \beta) = \mathbf{0} \text{ for } k \geq 2;$$

$$\widehat{\text{HL2}} \quad \mathbb{E} \left(\log^+ \rho(\boldsymbol{\varphi}_{(k),1}) \right) < \infty;$$

$$\widehat{\text{HL3}} \quad \mathbb{E} \left(\log \rho(\boldsymbol{\varphi}_{(k),t}^{(r)}) \right) < 0 \text{ for some integer } r \geq 1, \text{ where } \boldsymbol{\varphi}_{(k),t}^{(r)} = \boldsymbol{\varphi}_{(k),t} \circ \boldsymbol{\varphi}_{(k),t-1} \circ \dots \circ \boldsymbol{\varphi}_{(k),t-r+1} \text{ is the } r\text{-fold backward iterates of } \boldsymbol{\varphi}_{(k),t};$$

$$\widehat{\text{HL4}} \quad \mathbb{E} \left(\log^+ \|\mathbf{d}_0^{(k)}\|_{\boldsymbol{\Theta}_0^\epsilon \times \mathcal{B}} \right) < \infty, \text{ where } \{\mathbf{d}_t^{(k)}(\cdot), t \in \mathbb{Z}\} \text{ is the unique SE solution of the unperturbed system (with the existence guaranteed by Conditions } \widehat{\text{HL1}}\text{--}\widehat{\text{HL3}}\text{);}$$

$$\widehat{\text{HL5}} \quad \|\hat{\boldsymbol{\varphi}}_{(k),t}(\hat{\mathbf{f}}_1^{(k)}) - \boldsymbol{\varphi}_{(k),t}(\hat{\mathbf{f}}_1^{(k)})\|_{\boldsymbol{\Theta}_0^\epsilon \times \mathcal{B}} \xrightarrow{e.a.s.} 0 \text{ and } \rho(\hat{\boldsymbol{\varphi}}_{(k),t} - \boldsymbol{\varphi}_{(k),t}) \xrightarrow{e.a.s.} 0 \text{ as } t \rightarrow \infty.$$

Case $k = 1$: Under the condition that Assumption 2(ii) for $k = 0$ holds, the sequence $\{(\mathbf{y}_t, \mathbf{x}_t, f_t), t \in \mathbb{Z}\}$ is SE. Consequently, the sequence $\{\boldsymbol{\varphi}_{(1),t}, t \in \mathbb{Z}\}$ is SE as well. Condition $\widehat{\text{HL1}}$ follows directly from Assumption 5(i). Note that $\mathbf{d}^{(1)} \mapsto \boldsymbol{\varphi}_{(1),t}(\mathbf{d}^{(1)})$ is differentiable

everywhere on $\mathbb{R}^{m+2}$. Condition $\widehat{\text{HL3}}$ follows directly from the local contraction condition in [Assumption 5\(i\)](#) for $r = 1$, by applying a mean value theorem (see, e.g., [Rudin, 1976](#), Theorem 9.19). This, in turn, immediately implies Condition $\widehat{\text{HL2}}$.

Together, these three conditions ensure that the SE solution $\{\mathbf{d}_t^{(1)}, t \in \mathbb{Z}\}$ to the unperturbed system admits an almost sure representation ([Straumann and Mikosch, 2006](#), Eq. (2.5)):

$$\mathbf{d}_t^{(1)}(\boldsymbol{\vartheta}, \beta) = \sum_{j=0}^{\infty} \left(\prod_{\ell=1}^j Q_{t-\ell}(f_{t-\ell}(\boldsymbol{\vartheta}, \beta), \boldsymbol{\vartheta}, \beta) \right) \mathbf{q}_{(1), t-j-1}(f_{t-j-1}(\boldsymbol{\vartheta}, \beta), \boldsymbol{\vartheta}, \beta), \quad (\text{A.10})$$

with $\prod_{\ell=1}^0 \cdot \equiv 1$.

To establish Condition $\widehat{\text{HL4}}$, it suffices to show that $\left[\mathbb{E} \left(\|\mathbf{d}_t^{(1)}\|_{\boldsymbol{\Theta}_0^\epsilon \times \mathcal{B}}^\delta \right) \right]^{1/\delta} < \infty$ for some $\delta > 0$. Before proceeding, note that for a random sequence of possibly matrix-valued, dimension-compatible quantities $\mathbf{A}_j$ and $\mathbf{B}_j$, $j \geq 0$, an application of Minkowski's inequality followed by Hölder's inequality yields, for $1 \leq \rho_1 < \rho_2$,

$$\begin{aligned} \left(\mathbb{E} \left\| \sum_{j=0}^{\infty} \mathbf{A}_j \mathbf{B}_j \right\|^{\rho_1} \right)^{1/\rho_1} &\leq \sum_{j=0}^{\infty} \left(\mathbb{E} \|\mathbf{A}_j \mathbf{B}_j\|^{\rho_1} \right)^{1/\rho_1} \\ &\leq \sum_{j=0}^{\infty} \left(\mathbb{E} \|\mathbf{A}_j\|^{(\rho_1^{-1} - \rho_2^{-1})^{-1}} \right)^{\rho_1^{-1} - \rho_2^{-1}} \left(\mathbb{E} \|\mathbf{B}_j\|^{\rho_2} \right)^{\rho_2^{-1}} \\ &\leq C \sum_{j=0}^{\infty} \left(\mathbb{E} \|\mathbf{A}_j\|^{(\rho_1^{-1} - \rho_2^{-1})^{-1}} \right)^{\rho_1^{-1} - \rho_2^{-1}}, \end{aligned} \quad (\text{A.11})$$

whenever $\sup_{j \geq 0} \mathbb{E} \|\mathbf{B}_j\|^{\rho_2} < \infty$. Note that since $\{\mathbf{q}_{(1), t}(f_t(\boldsymbol{\vartheta}, \beta), \boldsymbol{\vartheta}, \beta), t \in \mathbb{Z}\}$ is strictly stationary, it follows that for all $t \in \mathbb{Z}$, $\mathbb{E} \left(\sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta}_0^\epsilon \times \mathcal{B}} \|\mathbf{q}_{(1), t}(f_t(\boldsymbol{\vartheta}, \beta), \boldsymbol{\vartheta}, \beta)\|^\eta \right) \leq C$ by [Assumption 5\(i\)](#). Recall the constants η and κ from [Assumption 5](#). If at least one is finite, set $\delta = (\eta^{-1} + \kappa^{-1})^{-1}$; if both are ∞ , δ is arbitrary subject to $\delta \geq 1$. We present the proofs for the case $\eta, \kappa < \infty$; if one is ∞ , only the first inequality in (A.11) is needed, after which the steps proceed straightforwardly. When $\eta, \kappa < \infty$, [Assumption 5](#) ensures that $\delta \in [1, \kappa \wedge \eta)$. Set $\rho_1 = \delta$ and $\rho_2 = \eta$ in (A.11) above and note that $1 \leq \rho_1 < \rho_2$. Condition $\widehat{\text{HL4}}$ then follows from

(A.11) and Assumption 5(ii). Specifically, for $\epsilon > 0$, by (A.10) and (A.11), we have

$$\begin{aligned}
& \left[\mathbb{E} \left(\left\| \mathbf{d}_t^{(1)} \right\|_{\Theta_0^\epsilon \times \mathcal{B}}^\delta \right) \right]^{1/\delta} \\
&= \left\{ \mathbb{E} \left(\sup_{(\boldsymbol{\vartheta}, \beta) \in \Theta_0^\epsilon \times \mathcal{B}} \left\| \sum_{j=0}^{\infty} \left(\prod_{\ell=1}^j Q_{t-\ell}(f_{t-\ell}(\boldsymbol{\vartheta}, \beta), \boldsymbol{\vartheta}, \beta) \right) \mathbf{q}_{(1), t-j-1}(f_{t-j-1}(\boldsymbol{\vartheta}, \beta), \boldsymbol{\vartheta}, \beta) \right\| \right)^\delta \right\}^{1/\delta} \\
&\leq C \sum_{j=0}^{\infty} \left\{ \mathbb{E} \left(\sup_{(\boldsymbol{\vartheta}, \beta) \in \Theta_0^\epsilon \times \mathcal{B}} \left| \prod_{\ell=1}^j \left(\beta + \alpha \frac{\partial s_{t-\ell}(f, \phi)}{\partial f} \Big|_{f=f_{t-\ell}(\boldsymbol{\vartheta}, \beta)} \right) \right|^\kappa \right) \right\}^{1/\kappa} \\
&\leq C \sum_{j=1}^{\infty} [\varrho_j(\boldsymbol{\vartheta}_0, \epsilon, \kappa)]^{1/\kappa} < \infty.
\end{aligned}$$

It remains to consider Condition $\widehat{\text{HL5}}$. Note that $\left\| \widehat{\boldsymbol{\varphi}}_{(1), t}(\hat{\mathbf{f}}_1^{(1)}) - \boldsymbol{\varphi}_{(1), t}(\hat{\mathbf{f}}_1^{(1)}) \right\|_{\Theta_0^\epsilon \times \mathcal{B}} \leq C(R_{1,t} + R_{2,t} + R_{3,t})$, where

$$R_{1,t} := \sup_{(\boldsymbol{\vartheta}, \beta) \in \Theta_0^\epsilon \times \mathcal{B}} \left| s_t(\hat{f}_t(\boldsymbol{\vartheta}, \beta), \phi) - s_t(f_t(\boldsymbol{\vartheta}, \beta), \phi) \right|, \quad (\text{A.12})$$

$$R_{2,t} := \sup_{(\boldsymbol{\vartheta}, \beta) \in \Theta_0^\epsilon \times \mathcal{B}} \left\| \frac{\partial s_t(f, \phi)}{\partial \phi} \Big|_{f=\hat{f}_t(\boldsymbol{\vartheta}, \beta)} - \frac{\partial s_t(f, \phi)}{\partial \phi} \Big|_{f=f_t(\boldsymbol{\vartheta}, \beta)} \right\|, \quad (\text{A.13})$$

$$R_{3,t} := \sup_{(\boldsymbol{\vartheta}, \beta) \in \Theta_0^\epsilon \times \mathcal{B}} \left| \frac{\partial s_t(f, \phi)}{\partial f} \Big|_{f=\hat{f}_t(\boldsymbol{\vartheta}, \beta)} - \frac{\partial s_t(f, \phi)}{\partial f} \Big|_{f=f_t(\boldsymbol{\vartheta}, \beta)} \right|. \quad (\text{A.14})$$

By (3.8) in Assumption 5(iii), we have $R_{1,t} \leq C_t(\boldsymbol{\vartheta}_0, \epsilon) \|\hat{f}_t(\boldsymbol{\vartheta}, \beta) - f_t(\boldsymbol{\vartheta}, \beta)\|_{\Theta_0^\epsilon \times \mathcal{B}}^\zeta$ for some $\zeta \in (0, 1]$, where $C_t(\boldsymbol{\vartheta}_0, \epsilon)$ is strictly stationary with $\mathbb{E}[\log^+ C_t(\boldsymbol{\vartheta}_0, \epsilon)] < \infty$. By Assumption 2(ii) for $k = 0$ and Lemma 2.1 of Straumann and Mikosch (2006), it follows that $R_{1,t} \xrightarrow{e.a.s.} 0$ as $t \rightarrow \infty$. Similarly, one can show that $R_{2,t} \xrightarrow{e.a.s.} 0$ and $R_{3,t} \xrightarrow{e.a.s.} 0$ as $t \rightarrow \infty$ as well. Finally, by applying (3.10) in Assumption 5(iii), we obtain $\rho(\widehat{\boldsymbol{\varphi}}_{(1), t} - \boldsymbol{\varphi}_{(1), t}) \leq C R_{3,t} \xrightarrow{e.a.s.} 0$. Given Conditions $\widehat{\text{HL1}}\text{--}\widehat{\text{HL5}}$, Theorem 2.10 of Straumann and Mikosch (2006) implies that $\|\hat{\mathbf{f}}_t^{(1)} - \mathbf{d}_t^{(1)}\|_{\Theta_0^\epsilon \times \mathcal{B}} \xrightarrow{e.a.s.} 0$ as $t \rightarrow \infty$. Moreover, $\mathbf{d}_t^{(1)}$ is $\mathcal{F}_{t-1}$ -measurable and has the almost sure representation $\mathbf{d}_t^{(1)} = \lim_{r \rightarrow \infty} \boldsymbol{\varphi}_{(1), t}^{(r)}(\hat{\mathbf{f}}_1^{(1)})$ for all $t \in \mathbb{Z}$, where the limit is independent of $\hat{f}_1$ and convergence is in the norm $\|\cdot\|_{\Theta_0^\epsilon \times \mathcal{B}}$. Note that for each $r \geq 1$, $\boldsymbol{\varphi}_{(1), t}^{(r)}(\hat{\mathbf{f}}_1^{(1)})$ is continuous on $\Theta_0^\epsilon \times \mathcal{B}$ by mathematical induction. Since $\Theta_0^\epsilon \times \mathcal{B}$ is compact, by the classical result that uniform limits of continuous functions are continuous Rudin (1976, Theorem 7.12), it follows that the limit function $\mathbf{d}_t^{(1)}(\cdot)$ is uniformly continuous on $\Theta_0^\epsilon \times \mathcal{B}$. Following the argument in Part (3) on p. 2483 of Straumann and Mikosch (2006), one can conclude that $\mathbf{f}_t^{(1)} \equiv \mathbf{d}_t^{(1)}$ for $t \in \mathbb{Z}$. By the discussion on verifying Condition $\widehat{\text{HL4}}$, we observe that $\mathbb{E}(\|\mathbf{f}_t^{(1)}\|_{\Theta_0^{\delta_1} \times \mathcal{B}}^{\delta_1}) < \infty$ for $\delta_1 = (\eta^{-1} + \kappa^{-1})^{-1}$ when $\eta < \infty$ or $\kappa < \infty$, and arbitrary $\delta_1 \geq 1$ when $\eta = \kappa = \infty$. This completes the proof for the case of $k = 1$.

Case $k = 2$: Since the proof is similar to the case of $k = 1$, we outline only the key differences below. Note that $\{\boldsymbol{\varphi}_{(2),t}, t \in \mathbb{Z}\}$ is SE by the joint SE property of $\{(\mathbf{y}_t, \mathbf{x}_t, f_t, \mathbf{f}_t^{(1)}), t \in \mathbb{Z}\}$, which follows from [Assumption 2\(ii\)](#) for $k = 0, 1$. By the construction of $\mathbf{q}_{(2),t}$ below the SRE [\(A.9\)](#), the verification of Condition [HL1](#) requires: For $t \in \mathbb{Z}$, (iii.1) $\mathbb{E}(\log^+ \sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta}_0^\epsilon \times \mathcal{B}} \|\mathbf{f}_t^{(1)}(\boldsymbol{\vartheta}, \beta)\|) < \infty$; (iii.2) $\mathbb{E}\left(\log^+ \sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta}_0^\epsilon \times \mathcal{B}} \left\| \frac{\partial^j s_t(f, \phi)}{\partial f^j} \right\|_{f=f_t(\boldsymbol{\vartheta}, \beta)}\right) < \infty$ with $j = 1, 2$;

$$(iii.3) \mathbb{E}\left(\log^+ \sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta}_0^\epsilon \times \mathcal{B}} \left\| \frac{\partial s_t(f, \phi)}{\partial \boldsymbol{\vartheta}} \right\|_{f=f_t(\boldsymbol{\vartheta}, \beta)}\right) < \infty;$$

$$(iii.4) \mathbb{E}\left(\log^+ \sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta}_0^\epsilon \times \mathcal{B}} \left\| \frac{\partial^2 s_t(f, \phi)}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}^\top} \right\|_{f=f_t(\boldsymbol{\vartheta}, \beta)}\right) < \infty;$$

$$(iii.5) \mathbb{E}\left(\log^+ \sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta}_0^\epsilon \times \mathcal{B}} \left\| \frac{\partial^2 s_t(f, \phi)}{\partial f \partial \boldsymbol{\vartheta}} \right\|_{f=f_t(\boldsymbol{\vartheta}, \beta)}\right) < \infty.$$

Note that (iii.1) holds because $\mathbb{E}(\|\mathbf{f}_t^{(1)}\|_{\boldsymbol{\Theta}_0^\epsilon \times \mathcal{B}}^{\delta_1}) < \infty$ as established for the case of $k = 1$. Condition (iii.3) follows from [Assumption 5\(i\)](#), and the remaining conditions are ensured by [Assumption 5\(iv\)](#). This verifies Condition [HL1](#). The verification of Conditions [HL2](#)–[HL3](#) proceeds as in the case of $k = 1$.

We now consider Condition [HL4](#). Under Conditions [HL1](#)–[HL3](#), and analogously to [\(A.10\)](#), one has $\mathbf{d}_t^{(2)}(\boldsymbol{\vartheta}, \beta) = \sum_{j=0}^{\infty} \left(\prod_{\ell=1}^j Q_{t-\ell}(f_{t-\ell}(\boldsymbol{\vartheta}, \beta), \boldsymbol{\vartheta}, \beta) \right) \mathbf{q}_{(2),t-j-1}(\mathbf{f}_{t-j-1}^{(1)}(\boldsymbol{\vartheta}, \beta), f_{t-j-1}(\boldsymbol{\vartheta}, \beta), \boldsymbol{\vartheta}, \beta)$. Consider the first case, where all first- and second-order l/r partial derivatives of $(f, \phi) \mapsto s_t(f, \phi)$, evaluated at $f = f_t(\boldsymbol{\vartheta}, \beta)$, are uniformly bounded over $(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta}_0^\epsilon \times \mathcal{B}$ (i.e., the first part in [Assumption 5\(iv\)](#) holds). By the definition of $\mathbf{q}_{(2),t}$ below [\(A.9\)](#), it is not hard to see $\mathbb{E}\left(\sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta}_0^\epsilon \times \mathcal{B}} \|\mathbf{q}_{(2),t}(\mathbf{f}_t^{(1)}(\boldsymbol{\vartheta}, \beta), f_t(\boldsymbol{\vartheta}, \beta), \boldsymbol{\vartheta}, \beta)\|^{\delta_1/2}\right) < \infty$. Let $\tilde{\delta} = (2\delta_1^{-1} + \kappa^{-1})^{-1}$. Note that if $\eta < \infty$ or $\kappa < \infty$, then $\tilde{\delta} = (2\delta_1^{-1} + 3\kappa^{-1})^{-1} \geq 1$ by [Assumption 5\(iv\)](#), and if $\eta = \kappa = \infty$, we have $\tilde{\delta} = \delta_1/2 \geq 1$ by choosing any $\delta_1 \geq 2$ (noting that in this case, δ_1 can be any value greater than or equal to 1). By setting $\rho_1 = \tilde{\delta}$ and $\rho_2 = \delta_1/2$ in [\(A.11\)](#), we obtain $\mathbb{E}(\|\mathbf{d}_t^{(2)}\|_{\boldsymbol{\Theta}_0^\epsilon \times \mathcal{B}}^{\tilde{\delta}}) < \infty$.

Consider the second case, where these partial derivatives are not necessarily bounded. For the moment, let $\lambda < \delta_1$ and note that $\|\text{vec}(\cdot)\| \leq \sqrt{\text{rank}(\cdot)} \|\cdot\|$. By employing the c_r -inequality and Hölder's inequality, we obtain $\mathbb{E}\left(\sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta}_0^\epsilon \times \mathcal{B}} \|\mathbf{q}_{(2),t}(\mathbf{f}_t^{(1)}(\boldsymbol{\vartheta}, \beta), f_t(\boldsymbol{\vartheta}, \beta), \boldsymbol{\vartheta}, \beta)\|^{\lambda/2}\right) < \infty$ provided that $\mathbb{E}(\|\mathbf{f}_t^{(1)}\|_{\boldsymbol{\Theta}_0^\epsilon \times \mathcal{B}}^{\delta_1}) < \infty$, $\mathbb{E}\left(\sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta}_0^\epsilon \times \mathcal{B}} \left\| \frac{\partial s_t(f, \phi)}{\partial f} \right\|_{f=f_t(\boldsymbol{\vartheta}, \beta)}^\lambda\right) < \infty$, $\mathbb{E}\left(\sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta}_0^\epsilon \times \mathcal{B}} \left\| \frac{\partial^2 s_t(f, \phi)}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}^\top} \right\|_{f=f_t(\boldsymbol{\vartheta}, \beta)}^{\lambda/2}\right) < \infty$, $\mathbb{E}\left(\sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta}_0^\epsilon \times \mathcal{B}} \left\| \frac{\partial^2 s_t(f, \phi)}{\partial f \partial \boldsymbol{\vartheta}} \right\|_{f=f_t(\boldsymbol{\vartheta}, \beta)}^\lambda\right) < \infty$,

and

$$\mathbb{E} \left(\sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta}_0^\epsilon \times \mathcal{B}} \left\| \frac{\partial s_t(f, \phi)}{\partial \boldsymbol{\vartheta}} \right|_{f=f_t(\boldsymbol{\vartheta}, \beta)} \right\|^{\lambda/2} \right) < \infty, \quad (\text{A.15})$$

$$\mathbb{E} \left(\sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta}_0^\epsilon \times \mathcal{B}} \left\| \frac{\partial^2 s_t(f, \phi)}{\partial f^2} \right|_{f=f_t(\boldsymbol{\vartheta}, \beta)} \right\|^{\lambda/2(1-\lambda/\delta_1)^{-1}} \right) < \infty. \quad (\text{A.16})$$

Since $\lambda/2 < \delta_1/2 < \eta$ (given $\lambda < \delta_1$ as previously assumed), the moment condition (A.15) is guaranteed by Assumption 5(i). If at least one of κ or η is finite, choose $\lambda \in (0, \delta_1/2] \subset (0, \infty)$; in this case, (A.16) is satisfied if

$$\mathbb{E} \left(\sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta}_0^\epsilon \times \mathcal{B}} \left\| \frac{\partial^2 s_t(f, \phi)}{\partial f^2} \right|_{f=f_t(\boldsymbol{\vartheta}, \beta)} \right\|^\lambda \right) < \infty. \quad (\text{A.17})$$

If $\tilde{\delta} = (2\lambda^{-1} + \kappa^{-1})^{-1} \geq 1$, we then obtain $\mathbb{E}(\|\mathbf{d}_t^{(2)}\|_{\boldsymbol{\Theta}_0^\epsilon \times \mathcal{B}}^{\tilde{\delta}}) < \infty$ using (A.11) with $\rho_1 = \tilde{\delta}$ and $\rho_2 = \lambda/2$, together with the moment conditions in Assumption 5(iv). If $\eta = \kappa = \infty$, then δ_1 can be any value greater than or equal to 1, and we can thus take $\delta_1 = 2\lambda$. In this case, (A.16) is equivalent to (A.17), and we again obtain $\mathbb{E}(\|\mathbf{d}_t^{(2)}\|_{\boldsymbol{\Theta}_0^\epsilon \times \mathcal{B}}^{\tilde{\delta}}) < \infty$ with Assumption 5(iv).

Finally, we address Condition $\widehat{\text{HL}}5$. Recall that $\mathbf{f}_1^{(2)}(\boldsymbol{\vartheta}, \beta) = \mathbf{0}$ for all $(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta} \times \mathcal{B}$, and thus we obtain

$$\begin{aligned} & \left\| \widehat{\boldsymbol{\varphi}}_{(2),t}(\hat{\mathbf{f}}_1^{(2)}) - \boldsymbol{\varphi}_{(2),t}(\hat{\mathbf{f}}_1^{(2)}) \right\|_{\boldsymbol{\Theta}_0^\epsilon \times \mathcal{B}} \\ &= \sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta}_0^\epsilon \times \mathcal{B}} \left\| \mathbf{q}_{(2),t}(\hat{\mathbf{f}}_t^{(1)}(\boldsymbol{\vartheta}, \beta), \hat{f}_t(\boldsymbol{\vartheta}, \beta), \boldsymbol{\vartheta}, \beta) - \mathbf{q}_{(2),t}(\mathbf{f}_t^{(1)}(\boldsymbol{\vartheta}, \beta), f_t(\boldsymbol{\vartheta}, \beta), \boldsymbol{\vartheta}, \beta) \right\|. \end{aligned}$$

By employing the inequality $\|\mathbf{C}_1 \mathbf{D}_1 - \mathbf{C}_2 \mathbf{D}_2\| \leq \|\mathbf{C}_1 - \mathbf{C}_2\| \|\mathbf{D}_2\| + \|\mathbf{D}_1 - \mathbf{D}_2\| \|\mathbf{C}_2\| + \|\mathbf{C}_1 - \mathbf{C}_2\| \|\mathbf{D}_1 - \mathbf{D}_2\|$ repeatedly, it follows that $\|\widehat{\boldsymbol{\varphi}}_{(2),t}(\hat{\mathbf{f}}_1^{(2)}) - \boldsymbol{\varphi}_{(2),t}(\hat{\mathbf{f}}_1^{(2)})\|_{\boldsymbol{\Theta}_0^\epsilon \times \mathcal{B}} \xrightarrow{e.a.s.} 0$ which is a consequence of the local Lipschitz properties in Assumption 5(iii), combined with the moment conditions in Assumption 5(iv) and Lemma 2.1 of Straumann and Mikosch (2006). Finally, $\rho(\widehat{\boldsymbol{\varphi}}_{(2),t} - \boldsymbol{\varphi}_{(2),t}) \xrightarrow{e.a.s.} 0$ follows directly from the same reasoning as in the case of $k = 1$. Similarly, $\mathbf{d}_t^{(2)}(\cdot)$ is uniformly continuous on $\boldsymbol{\Theta}_0^\epsilon \times \mathcal{B}$, and following the argument in Part (3) on p. 2483 of Straumann and Mikosch (2006), we conclude that $\mathbf{f}_t^{(2)} \equiv \mathbf{d}_t^{(2)}$ for $t \in \mathbb{Z}$. Based on the discussion for verifying Condition $\widehat{\text{HL}}4$, we further conclude that $\mathbb{E}(\|\mathbf{f}_t^{(2)}\|_{\boldsymbol{\Theta}_0^\epsilon \times \mathcal{B}}^{\delta_2}) < \infty$, where $\delta_2 = (2\delta_1^{-1} + \kappa^{-1})^{-1}$ the uniform boundedness conditions in Assumption 5(iv) hold, and $\delta_2 = (2\lambda^{-1} + \kappa^{-1})^{-1}$ otherwise. This completes the proof for the case $k = 2$. $\square$

Lemma A.1. Suppose Assumptions 1 and 2(i) hold, and the following conditions are satisfied:

$$IV1 \quad \mathbb{E}(\log^+ \sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta} \times \mathcal{B}} |s(\mathbf{y}_t, \mathbf{x}_t, \hat{f}_1(\boldsymbol{\vartheta}, \beta), \phi)|) < \infty \text{ for } \hat{f}_1(\boldsymbol{\vartheta}, \beta) = \omega \in \Theta_\omega.$$

$$IV2 \quad \mathbb{E}\left(\log^+ \sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta} \times \mathcal{B}} \sup_{f \in \mathcal{F}} \left| \beta + \alpha \frac{\partial}{\partial f} s(\mathbf{y}_t, \mathbf{x}_t, f, \phi) \right| \right) < \infty.$$

$$IV3 \quad \mathbb{E}\left(\log \sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta} \times \mathcal{B}} \sup_{f \in \mathcal{F}} \left| \frac{\partial}{\partial f} \psi_t^{(r)}(f, \boldsymbol{\vartheta}, \beta) \right| \right) < 0 \text{ for some } r \in \mathbb{Z}^+, \text{ where } \psi_t^{(r)}(\cdot, \boldsymbol{\vartheta}, \beta) := \psi_t(\cdot, \boldsymbol{\vartheta}, \beta) \circ \psi_{t-1}(\cdot, \boldsymbol{\vartheta}, \beta) \circ \dots \circ \psi_{t-r+1}(\cdot, \boldsymbol{\vartheta}, \beta) \text{ is the } r\text{-fold backward iterates of } \psi_t(\cdot, \boldsymbol{\vartheta}, \beta).$$

Then Assumption 2(ii) holds for $k = 0$.

Proof of Lemma A.1. Note that for all $t \in \mathbb{Z}$, $s_t(f, \phi)$ is continuous in $(f, \phi) \in \mathcal{F} \times \boldsymbol{\Theta}_\phi$ under Assumption 1(iii). By mathematical induction, it follows that for $t \in \mathbb{Z}^+$, the initialized sequence $\hat{f}_t(\boldsymbol{\vartheta}, \beta)$, with the initial value $\hat{f}_1(\boldsymbol{\vartheta}, \beta) = \omega$, is continuous in $(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta} \times \mathcal{B}$. Since $\boldsymbol{\Theta} \times \mathcal{B}$ is compact, $\hat{f}_t(\cdot)$ is also uniformly continuous. We can treat the sequence $\{\hat{f}_t(\cdot), t \in \mathbb{Z}^+\}$ as a sequence of random elements in the space $(\mathcal{C}^0(\boldsymbol{\Theta} \times \mathcal{B}, \mathcal{F}), \|\cdot\|_{\boldsymbol{\Theta} \times \mathcal{B}})$. One can write $\hat{f}_{t+1} = \varphi_t(\hat{f}_t)$, where the random maps $\varphi_t : (\mathcal{C}^0(\boldsymbol{\Theta} \times \mathcal{B}, \mathcal{F}), \|\cdot\|_{\boldsymbol{\Theta} \times \mathcal{B}}) \rightarrow (\mathcal{C}^0(\boldsymbol{\Theta} \times \mathcal{B}, \mathcal{F}), \|\cdot\|_{\boldsymbol{\Theta} \times \mathcal{B}})$ are given by $[\varphi_t(f)](\boldsymbol{\vartheta}, \beta) = \psi_t(f(\boldsymbol{\vartheta}, \beta), \boldsymbol{\vartheta}, \beta)$ with $\psi_t(\cdot, \boldsymbol{\vartheta}, \beta)$ provided in (2.1). Note that $\{\varphi_t, t \in \mathbb{Z}\}$ is SE under Assumption 2(i).

The following high-level conditions suffice to apply Theorem 3.1 of Bougerol (1993) and Theorem 2.8 of Straumann and Mikosch (2006): (i.1) $\mathbb{E}(\log^+ \|\varphi_1(\hat{f}_1) - \hat{f}_1\|_{\boldsymbol{\Theta} \times \mathcal{B}}) < \infty$, where $\hat{f}_1(\boldsymbol{\vartheta}, \beta) = \omega$ for $(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta} \times \mathcal{B}$; (i.2) $\mathbb{E}(\log^+ \rho(\varphi_1)) < \infty$; (i.3) $\mathbb{E}(\log \rho(\varphi_t^{(r)})) < 0$ for some integer $r \geq 1$, where $\varphi_t^{(r)} = \varphi_t \circ \varphi_{t-1} \circ \dots \circ \varphi_{t-r+1}$ is the r -fold convolution (backward iterates) of φ_t . These three high-level conditions can be readily verified using Conditions IV1–IV3, following standard arguments; see, e.g., Blasques et al. (2022, Proposition 3.2) or Lin and Lucas (2025, Proposition 3). Applying Straumann and Mikosch (2006, Theorem 2.8), we conclude that the stochastic recurrence equation (SRE) $h_{t+1} = \varphi_t(h_t)$, $t \in \mathbb{Z}$, admits a unique SE solution $\{f_t, t \in \mathbb{Z}\}$ such that $\|\hat{f}_t - f_t\|_{\boldsymbol{\Theta} \times \mathcal{B}} = \sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta} \times \mathcal{B}} |\hat{f}_t(\boldsymbol{\vartheta}, \beta) - f_t(\boldsymbol{\vartheta}, \beta)| \xrightarrow{e.a.s.} 0$ as $t \rightarrow \infty$. As $\boldsymbol{\Theta} \times \mathcal{B}$ is compact, applying the same argument as for the case of $k = 1$ in Proposition 1, the uniform continuity of $f_t(\cdot)$ immediately follows. This establishes Lemma A.1. $\square$

Lemma A.2 (Identifiable uniqueness). Suppose Assumptions 1, 2, and 3(i) are satisfied. Under correct model specification, if $p(\mathbf{y} | \mathbf{x}, f, \phi) = p(\mathbf{y} | \mathbf{x}, \tilde{f}, \tilde{\phi})$ for almost every $(\mathbf{y}, \mathbf{x}) \in \mathcal{Y} \times \mathcal{X}$ (with respect to an appropriate dominating measure on $\mathcal{Y} \times \mathcal{X}$) if and only if $f = \tilde{f}$ and $\phi = \tilde{\phi}$, then Assumption 3(ii) holds.

Proof of Lemma A.2. As noted below (A.1) in the proof of Theorem 1, three important implications follow: first, the sequence $\{\ell_t(f_t(\boldsymbol{\vartheta}, \beta), \boldsymbol{\phi}), t \in \mathbb{Z}\}$ is SE; second, defining $\mathcal{L}(\boldsymbol{\vartheta}, \beta) := \mathbb{E}[\ell_t(f_t(\boldsymbol{\vartheta}, \beta), \boldsymbol{\phi})]$, it is established there that $\mathcal{L}$ is continuous on the compact set $\boldsymbol{\Theta} \times \mathcal{B}$; and third, $\mathcal{L}^*(\boldsymbol{\vartheta}_0) := \mathcal{L}(\boldsymbol{\vartheta}_0, \beta)$ is independent of $\beta \in \mathcal{B}$. Given that $\mathcal{B}$ is compact and $\mathcal{L}$ is continuous, it suffices to show that $\mathcal{L}(\boldsymbol{\vartheta}, \beta)$ is uniquely maximized over $\boldsymbol{\Theta}$ at $\boldsymbol{\vartheta}_0$ for each $\beta \in \mathcal{B}$. By Gibbs' inequality for the Kullback-Leibler divergence (White, 1996, Theorem 2.3), it follows that for continuous random variables $\mathbf{y}_t$:

$$\mathcal{L}^*(\boldsymbol{\vartheta}_0) - \mathcal{L}(\boldsymbol{\vartheta}, \beta) = \mathbb{E} \left[\int p(\mathbf{y}_t \mid \mathbf{x}_t, f_t(\boldsymbol{\vartheta}_0, \beta), \boldsymbol{\phi}_0) \log \frac{p(\mathbf{y}_t \mid \mathbf{x}_t, f_t(\boldsymbol{\vartheta}_0, \beta), \boldsymbol{\phi}_0)}{p(\mathbf{y}_t \mid \mathbf{x}_t, f_t(\boldsymbol{\vartheta}, \beta), \boldsymbol{\phi})} d\mathbf{y}_t \right] \geq 0,$$

with strict inequality unless $p(\mathbf{y}_t \mid \mathbf{x}_t, f_t(\boldsymbol{\vartheta}_0, \beta), \boldsymbol{\phi}_0) = p(\mathbf{y}_t \mid \mathbf{x}_t, f_t(\boldsymbol{\vartheta}, \beta), \boldsymbol{\phi})$. That is, under the condition above, $\mathcal{L}^*(\boldsymbol{\vartheta}_0) = \mathcal{L}(\boldsymbol{\vartheta}, \beta)$ implies that $f_t(\boldsymbol{\vartheta}, \beta) = f_t(\boldsymbol{\vartheta}_0, \beta)$ a.s. for all $t \in \mathbb{Z}$ and $\boldsymbol{\phi} = \boldsymbol{\phi}_0$. We arrive at $0 = (\omega - \omega_0)(1 - \beta) + \alpha s_t(f_t(\boldsymbol{\vartheta}, \beta), \boldsymbol{\phi})$ a.s. Since $s_t(f_t(\boldsymbol{\vartheta}, \beta), \boldsymbol{\phi})$ is random, one must have $\alpha = 0 = \alpha_0$, which further leads to $\omega = \omega_0$. That is, $\mathcal{L}^*(\boldsymbol{\vartheta}_0) = \mathcal{L}(\boldsymbol{\vartheta}, \beta)$ if and only if $\boldsymbol{\vartheta} = \boldsymbol{\vartheta}_0$. If $\mathbf{y}_t$ are discrete, the discrete version of Gibbs' inequality applies instead. $\square$

B Simulated critical values

We present the simulated critical values for the limiting distribution in (3.7), with results reported in Table B.3. To obtain these values, we approximate the infinite sum $(1 - \beta^2)^{1/2} \sum_{j=0}^{\infty} \beta^j X_j$ by $X(\beta) := (1 - \beta^2)^{1/2} \sum_{j=0}^{J_{\max}} \beta^j X_j$ for some $J_{\max} \in \mathbb{Z}^+$. We then maximize over $\beta \in [\beta_L, \beta_U]$ the quantity $(\max\{X(\beta), 0\})^2$ for $\alpha_L = 0$ and $[X(\beta)]^2$ for $\alpha_L < 0$, respectively. The results are based on 10^5 simulations with $J_{\max} = 3 \cdot 10^4$.

Table B.3: Simulated critical values for the limiting distribution in (3.7), with $\Theta_\alpha = [\alpha_L, \alpha_U]$ for $\alpha_U > 0$ and $\mathcal{B} = [\beta_L, \beta_U]$.

β_L	β_U	$\alpha_L = 0$			$\alpha_L < 0$			β_L	β_U	$\alpha_L = 0$			$\alpha_L < 0$		
		10%	5%	1%	10%	5%	1%			10%	5%	1%	10%	5%	1%
0	0.995	3.365	4.719	7.855	4.437	5.850	9.127	-0.995	0.995	3.913	5.362	8.714	5.014	6.507	9.909
0	0.990	3.266	4.613	7.696	4.416	5.816	9.048	-0.990	0.990	3.842	5.278	8.629	4.996	6.481	9.886
0	0.980	3.166	4.502	7.540	4.371	5.753	8.962	-0.980	0.980	3.760	5.191	8.506	4.964	6.442	9.814
0	0.970	3.092	4.417	7.425	4.322	5.691	8.854	-0.970	0.970	3.703	5.128	8.423	4.936	6.405	9.746
0	0.960	3.037	4.355	7.363	4.278	5.638	8.794	-0.960	0.960	3.657	5.076	8.343	4.905	6.376	9.706
0	0.950	2.989	4.308	7.277	4.247	5.587	8.731	-0.950	0.950	3.614	5.021	8.298	4.873	6.336	9.673
0	0.900	2.810	4.106	7.080	4.084	5.403	8.502	-0.900	0.900	3.456	4.840	8.116	4.748	6.155	9.470
0	0.850	2.675	3.949	6.930	3.958	5.236	8.300	-0.850	0.850	3.319	4.688	7.909	4.626	6.000	9.267
0	0.800	2.568	3.831	6.813	3.840	5.114	8.169	-0.800	0.800	3.197	4.545	7.699	4.499	5.857	9.100
0	0.750	2.490	3.732	6.720	3.737	5.007	8.043	-0.750	0.750	3.087	4.411	7.546	4.382	5.728	8.919
0	0.700	2.417	3.644	6.619	3.651	4.908	7.955	-0.700	0.700	2.981	4.300	7.378	4.271	5.614	8.767
0	0.650	2.341	3.567	6.530	3.568	4.805	7.862	-0.650	0.650	2.880	4.183	7.265	4.165	5.483	8.598
0	0.600	2.272	3.488	6.451	3.489	4.734	7.730	-0.600	0.600	2.778	4.058	7.133	4.052	5.367	8.427
0	0.550	2.218	3.413	6.385	3.411	4.656	7.640	-0.550	0.550	2.689	3.944	6.981	3.954	5.226	8.296
0	0.500	2.158	3.339	6.296	3.342	4.579	7.529	-0.500	0.500	2.592	3.829	6.853	3.841	5.109	8.145
0	0.450	2.099	3.264	6.204	3.273	4.499	7.425	-0.450	0.450	2.499	3.731	6.722	3.736	4.987	8.009
0	0.400	2.042	3.207	6.102	3.210	4.419	7.335	-0.400	0.400	2.402	3.631	6.632	3.632	4.872	7.887
0	0.350	1.988	3.153	6.028	3.146	4.359	7.245	-0.350	0.350	2.307	3.524	6.529	3.522	4.763	7.745
0	0.300	1.932	3.094	5.978	3.090	4.275	7.166	-0.300	0.300	2.214	3.410	6.349	3.407	4.645	7.585
0	0.250	1.884	3.025	5.898	3.024	4.205	7.095	-0.250	0.250	2.120	3.297	6.232	3.301	4.521	7.432
0	0.200	1.835	2.969	5.811	2.968	4.135	6.999	-0.200	0.200	2.026	3.188	6.090	3.189	4.400	7.295
0.1	0.995	3.334	4.679	7.784	4.421	5.821	9.086	-0.1	0.995	3.403	4.757	7.896	4.463	5.868	9.185
0.1	0.990	3.226	4.572	7.619	4.395	5.788	8.997	-0.1	0.990	3.308	4.655	7.751	4.443	5.845	9.100
0.1	0.980	3.122	4.452	7.457	4.338	5.710	8.893	-0.1	0.980	3.206	4.537	7.609	4.404	5.792	8.997

continued on next page

β_L	β_U	$\alpha_L = 0$			$\alpha_L < 0$			β_L	β_U	$\alpha_L = 0$			$\alpha_L < 0$		
		10%	5%	1%	10%	5%	1%			10%	5%	1%	10%	5%	1%
0.1	0.970	3.044	4.367	7.377	4.288	5.652	8.789	-0.1	0.970	3.133	4.467	7.507	4.363	5.735	8.914
0.1	0.960	2.983	4.304	7.290	4.243	5.598	8.740	-0.1	0.960	3.086	4.413	7.429	4.331	5.687	8.844
0.1	0.950	2.937	4.245	7.222	4.204	5.541	8.670	-0.1	0.950	3.037	4.366	7.380	4.300	5.644	8.799
0.1	0.900	2.749	4.041	6.983	4.030	5.329	8.423	-0.1	0.900	2.867	4.173	7.176	4.146	5.469	8.604
0.1	0.850	2.614	3.877	6.842	3.887	5.154	8.230	-0.1	0.850	2.743	4.032	7.045	4.029	5.316	8.424
0.1	0.800	2.504	3.757	6.707	3.759	5.025	8.063	-0.1	0.800	2.643	3.913	6.900	3.920	5.193	8.287
0.3	0.995	3.265	4.594	7.702	4.370	5.770	8.944	-0.3	0.995	3.475	4.840	8.018	4.528	5.945	9.264
0.3	0.990	3.156	4.483	7.510	4.333	5.713	8.861	-0.3	0.990	3.389	4.754	7.855	4.513	5.926	9.231
0.3	0.980	3.034	4.348	7.331	4.264	5.619	8.746	-0.3	0.980	3.300	4.642	7.734	4.483	5.881	9.165
0.3	0.970	2.939	4.252	7.236	4.199	5.547	8.664	-0.3	0.970	3.230	4.569	7.649	4.449	5.831	9.072
0.3	0.960	2.878	4.173	7.154	4.143	5.479	8.569	-0.3	0.960	3.186	4.513	7.569	4.418	5.793	8.996
0.3	0.950	2.816	4.104	7.055	4.090	5.416	8.486	-0.3	0.950	3.144	4.477	7.508	4.387	5.760	8.944
0.3	0.900	2.615	3.876	6.798	3.883	5.149	8.230	-0.3	0.900	2.994	4.311	7.335	4.266	5.606	8.763
0.3	0.850	2.468	3.708	6.644	3.711	4.969	8.022	-0.3	0.850	2.885	4.181	7.215	4.163	5.482	8.616
0.3	0.800	2.351	3.582	6.491	3.581	4.820	7.843	-0.3	0.800	2.792	4.064	7.101	4.057	5.376	8.500
0.5	0.995	3.177	4.507	7.586	4.324	5.683	8.872	-0.5	0.995	3.564	4.918	8.140	4.602	6.016	9.352
0.5	0.990	3.056	4.385	7.389	4.261	5.604	8.744	-0.5	0.990	3.486	4.844	8.024	4.591	6.003	9.318
0.5	0.980	2.916	4.218	7.183	4.170	5.485	8.608	-0.5	0.980	3.403	4.748	7.892	4.565	5.965	9.273
0.5	0.970	2.817	4.100	7.057	4.084	5.390	8.463	-0.5	0.970	3.342	4.684	7.804	4.539	5.929	9.236
0.5	0.960	2.742	4.013	6.962	4.013	5.312	8.346	-0.5	0.960	3.298	4.641	7.746	4.516	5.893	9.190
0.5	0.950	2.677	3.942	6.869	3.946	5.227	8.261	-0.5	0.950	3.253	4.602	7.695	4.493	5.865	9.139
0.5	0.900	2.445	3.674	6.559	3.686	4.933	8.012	-0.5	0.900	3.120	4.444	7.519	4.392	5.737	8.947
0.5	0.850	2.276	3.484	6.343	3.494	4.729	7.735	-0.5	0.850	3.017	4.335	7.390	4.304	5.631	8.804
0.5	0.800	2.155	3.333	6.169	3.340	4.553	7.533	-0.5	0.800	2.932	4.242	7.266	4.212	5.541	8.679
0.7	0.995	3.056	4.377	7.431	4.245	5.618	8.798	-0.7	0.995	3.671	5.052	8.279	4.721	6.137	9.495
0.7	0.990	2.909	4.212	7.235	4.151	5.496	8.645	-0.7	0.990	3.603	4.968	8.182	4.710	6.128	9.467
0.7	0.980	2.742	4.015	6.971	4.015	5.319	8.398	-0.7	0.980	3.522	4.878	8.082	4.689	6.102	9.429
0.7	0.970	2.632	3.885	6.799	3.899	5.176	8.259	-0.7	0.970	3.467	4.827	8.012	4.666	6.068	9.383
0.7	0.960	2.534	3.773	6.689	3.791	5.057	8.131	-0.7	0.960	3.424	4.792	7.946	4.646	6.039	9.360
0.7	0.950	2.452	3.677	6.589	3.701	4.956	8.011	-0.7	0.950	3.388	4.755	7.878	4.630	6.020	9.334
0.7	0.900	2.191	3.375	6.205	3.381	4.595	7.637	-0.7	0.900	3.268	4.621	7.732	4.543	5.905	9.196
0.7	0.850	2.005	3.150	5.937	3.156	4.355	7.284	-0.7	0.850	3.176	4.506	7.614	4.465	5.819	9.044
0.7	0.800	1.855	2.962	5.708	2.975	4.155	7.036	-0.7	0.800	3.102	4.429	7.519	4.391	5.743	8.939

C Additional details on the four examples

We provide further details on the examples in Section 4.

C.1 Gaussian location model

For $t = 1, \dots, T$, consider the model $y_t = f_t + u_t$, where $u_t = \sigma_u \epsilon_t$ and ϵ_t are i.i.d. standard normal random variables. Assume that the sequence $\{y_t, t \in \mathbb{Z}\}$ is generated under the null hypothesis, i.e., $y_t \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(\omega_0, \sigma_{u0}^2)$. Under this assumption, the model is correctly specified. Note that $\nabla_t^f(f, \phi) = (y_t - f)/\sigma_u^2$, where $\phi = \sigma_u > 0$. Take $S_t(f, \phi) = \sigma_u^2$, then $s_t(f, \phi) = y_t - f$ and $f_{t+1} = \omega(1 - \beta) + \beta f_t + \alpha(y_t - f_t)$. We define $\mathcal{B} = [0, \beta_U]$ with $\beta_U < 1$, and $\Theta = \{\boldsymbol{\vartheta} = (\alpha, \omega, \phi)^\top \in \mathbb{R}^3 : 0 \leq \alpha \leq \alpha_U < 1, \omega_L \leq \omega \leq \omega_U, 0 < \sigma_L \leq \phi \leq \sigma_U\}$.

Let $\mathcal{F} = \mathbb{R}$, so that [Assumption 1](#) trivially holds for any $\epsilon > 0$. For [Assumption 2](#), [Assumption 2\(i\)](#) holds by assumption. We now verify the sufficient conditions in [Lemma A.1](#) to establish [Assumption 2\(ii\)](#) for $k = 0$. Condition [IV1](#) holds because $\mathbb{E} \log^+ |y_t| < \infty$ and $\omega_U < \infty$. Note that $\sup_{\boldsymbol{\vartheta} \in \Theta} |\beta - \alpha| = \alpha_U \vee \beta_U < 1$. We see that Condition [IV3](#) holds with $r = 1$, which implies that Condition [IV2](#) also holds. To verify [Assumption 2\(ii\)](#) for $k = 1, 2$, we first consider the sufficient conditions in [Assumption 5](#) and then apply [Proposition 1](#). We set $\eta = 4$ and $\kappa = \infty$. For [Assumption 5\(i\)](#), it is sufficient to show that

$$\mathbb{E} \left(\sup_{(\boldsymbol{\vartheta}, \beta) \in \Theta \times \mathcal{B}} |y_t - f_t(\boldsymbol{\vartheta}, \beta)| \right)^4 < \infty. \quad (\text{C.1})$$

We have $\mathbb{E} \left(\sup_{(\boldsymbol{\vartheta}, \beta) \in \Theta \times \mathcal{B}} |y_t - f_t(\boldsymbol{\vartheta}, \beta)| \right)^4 \leq 8 \mathbb{E}(y_t^4) + 8 \mathbb{E} \left(\sup_{(\boldsymbol{\vartheta}, \beta) \in \Theta \times \mathcal{B}} |f_t(\boldsymbol{\vartheta}, \beta)| \right)^4$ by the c_r -inequality. Since $f_t(\boldsymbol{\vartheta}, \beta)$ admits an a.s. representation $f_t(\boldsymbol{\vartheta}, \beta) = \sum_{j=0}^{\infty} (\beta - \alpha)^j [\omega(1 - \beta) + \alpha y_{t-1-j}]$, $t \in \mathbb{Z}$, we have $\sup_{(\boldsymbol{\vartheta}, \beta) \in \Theta \times \mathcal{B}} |f_t(\boldsymbol{\vartheta}, \beta)| \leq \sum_{j=0}^{\infty} (\alpha_U \vee \beta_U)^j (|\omega_L| \vee |\omega_U| + \alpha_U |y_{t-1-j}|)$.

By applying Minkowski's inequality, we arrive at

$$\begin{aligned} \left(\mathbb{E} \left(\sup_{(\boldsymbol{\vartheta}, \beta) \in \Theta \times \mathcal{B}} |f_t(\boldsymbol{\vartheta}, \beta)| \right)^4 \right)^{1/4} &\leq \sum_{j=0}^{\infty} \left\{ \mathbb{E} \left[(\alpha_U \vee \beta_U)^j (|\omega_L| \vee |\omega_U| + \alpha_U |y_{t-1-j}|) \right]^4 \right\}^{1/4} \\ &= \left(\mathbb{E} (|\omega_L| \vee |\omega_U| + \alpha_U |y_1|)^4 \right)^{1/4} (1 - \alpha_U \vee \beta_U)^{-1} < \infty. \end{aligned}$$

Then $\mathbb{E} \left(\sup_{(\boldsymbol{\vartheta}, \beta) \in \Theta \times \mathcal{B}} |y_t - f_t(\boldsymbol{\vartheta}, \beta)| \right)^4 < \infty$ follows from $\mathbb{E} \left(\sup_{(\boldsymbol{\vartheta}, \beta) \in \Theta \times \mathcal{B}} |f_t(\boldsymbol{\vartheta}, \beta)| \right)^4 < \infty$ and

$\mathbb{E}(y_t^4) < \infty$, thereby establishing [Assumption 5\(i\)](#). [Assumption 5\(ii\)](#) holds because

$$\sum_{j=1}^{\infty} \sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta}_0^\epsilon \times \mathcal{B}} \left| \prod_{\ell=1}^j (\beta - \alpha) \right| \leq \sum_{j=1}^{\infty} (\alpha_U \vee \beta_U)^j < \infty. \quad (\text{C.2})$$

[Assumption 5\(iii\)](#) holds by setting the sequence $C_t(\boldsymbol{\vartheta}_0, \epsilon) = 1$ for all $\ell = 0, 1, \dots, 5$ and $t \in \mathbb{Z}$. Moreover, [Assumption 5\(iv\)](#) holds since the uniform boundedness condition applies, and thus [Proposition 1](#) follows.

We now proceed to verify the conditions in [Assumption 3](#). Note that

$$\begin{aligned} \sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta} \times \mathcal{B}} |d_t(\boldsymbol{\vartheta}, \beta)| &\leq \frac{1}{2} \sigma_L^{-2} \left(\sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta} \times \mathcal{B}} |\hat{f}_t(\boldsymbol{\vartheta}, \beta) - f_t(\boldsymbol{\vartheta}, \beta)| \right)^2 \\ &\quad + \sigma_L^{-2} \sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta} \times \mathcal{B}} |y_t - f_t(\boldsymbol{\vartheta}, \beta)| \sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta} \times \mathcal{B}} |\hat{f}_t(\boldsymbol{\vartheta}, \beta) - f_t(\boldsymbol{\vartheta}, \beta)|. \end{aligned}$$

By [Lemma A.1](#) and Lemma 2.1 of [Straumann and Mikosch \(2006\)](#), $\sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta} \times \mathcal{B}} |d_t(\boldsymbol{\vartheta}, \beta)| \xrightarrow{e.a.s.} 0$ provided that $\mathbb{E} \left(\log^+ \sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta} \times \mathcal{B}} |y_t - f_t(\boldsymbol{\vartheta}, \beta)| \right) < \infty$. This moment condition is satisfied by [\(C.1\)](#). Moreover, [\(C.1\)](#) also ensures that $\mathbb{E} \left(\sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta} \times \mathcal{B}} |\ell_t(f_t(\boldsymbol{\vartheta}, \beta), \boldsymbol{\phi})| \right) < \infty$. [Assumption 3\(i\)](#) is obtained. [Assumption 3\(ii\)](#) follows directly from [Lemma A.2](#) under correct model specification. Next, consider [Assumption 3\(iii\)](#). By simple linear algebra, we have

$$\begin{aligned} \left\| \frac{\partial^2}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}^\top} \ell_t(f_t(\boldsymbol{\vartheta}, \beta), \boldsymbol{\phi}) \right\| &\leq \sigma_u^{-2} \|\mathbf{f}_t^{(1)}(\boldsymbol{\vartheta}, \beta)\|^2 + \left| \sigma_u^{-2} - 3\sigma_u^{-4} (y_t - f_t(\boldsymbol{\vartheta}, \beta)) \right| \\ &\quad + 4\sigma_u^{-3} |y_t - f_t(\boldsymbol{\vartheta}, \beta)| \|\mathbf{f}_t^{(1)}(\boldsymbol{\vartheta}, \beta)\| + \sigma_u^{-2} |y_t - f_t(\boldsymbol{\vartheta}, \beta)| \left\| \frac{\partial^2 f_t(\boldsymbol{\vartheta}, \beta)}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}^\top} \right\|. \end{aligned}$$

Since the uniform boundedness condition in [Assumption 5\(iv\)](#) holds, by [Proposition 1](#), we have $\mathbb{E} \left(\sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta}_0^\epsilon \times \mathcal{B}} \|\mathbf{f}_t^{(1)}(\boldsymbol{\vartheta}, \beta)\|^4 \right) < \infty$ and $\mathbb{E} \left(\sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta}_0^\epsilon \times \mathcal{B}} \|\mathbf{f}_t^{(2)}(\boldsymbol{\vartheta}, \beta)\|^2 \right) < \infty$. Combining these results with [\(C.1\)](#) and applying the Cauchy–Schwarz inequality, it follows that $\mathbb{E} \left(\sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta}_0^\epsilon \times \mathcal{B}} \left\| \frac{\partial^2}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}^\top} \ell_t(f_t(\boldsymbol{\vartheta}, \beta), \boldsymbol{\phi}) \right\| \right) < \infty$, and therefore [Assumption 3\(iii\)](#) holds. For [Assumption 3\(iv\)](#), we have $\nabla_t^\phi(f, \boldsymbol{\phi}) = -\sigma_u^{-1} + \sigma_u^{-3} (y_t - f)^2$, which gives $\nabla_t^\phi(\omega_0, \boldsymbol{\phi}_0) = \sigma_{u0}^{-1} (\epsilon_t^2 - 1)$. Therefore, $\mathbb{E}(\nabla_t^\phi(\omega_0, \boldsymbol{\phi}_0) \mid \mathcal{F}_{t-1}) = 0$ and $\mathbb{E}[(\nabla_t^\phi(\omega_0, \boldsymbol{\phi}_0))^2] = 2$. Moreover, we have $\mathbb{E}(s_t^2(\omega_0, \boldsymbol{\phi}_0)) = \sigma_{u0}^2$, $\mathbb{E}(\nabla_t^f(\omega_0, \boldsymbol{\phi}_0) \mid \mathcal{F}_{t-1}) = \sigma_{u0}^{-1} \mathbb{E}(\epsilon_t) = 0$, $\Sigma_{ff} = \mathbb{E}([\nabla_t^f(\omega_0, \boldsymbol{\phi}_0)]^2 \mid \mathcal{F}_{t-1}) = \sigma_{u0}^{-2}$, $\Omega_{ff} = -\mathbb{E}(\nabla_t^{ff}(\omega_0, \boldsymbol{\phi}_0) \mid \mathcal{F}_{t-1}) = \sigma_{u0}^{-2} = \Sigma_{ff}$, $\mathbb{E}(\nabla_t^\phi(\omega_0, \boldsymbol{\phi}_0) \nabla_t^f(\omega_0, \boldsymbol{\phi}_0) \mid \mathcal{F}_{t-1}) = \sigma_{u0}^{-2} \mathbb{E}(\epsilon_t^3 - \epsilon_t) = 0$. Since $\nabla_t^{\phi f}(f, \boldsymbol{\phi}) = -2\sigma_u^{-3} (y_t - f)$, one has $\Omega_{\phi f} = \mathbb{E}(\nabla_t^{\phi f}(\omega_0, \boldsymbol{\phi}_0) \mid \mathcal{F}_{t-1}) = 0$. All the conditions in [Assumption 3\(iv\)](#) are therefore satisfied. For [Assumption 3\(v\)](#), $\text{Var} \left(\sum_{j=0}^{\infty} \beta^j s_{t-j}(\omega_0, \boldsymbol{\phi}_0) \right) = \text{Var} \left(\sigma_{u0} \sum_{j=0}^{\infty} \beta^j \epsilon_{t-j} \right) = \sigma_{u0}^2 (1 - \beta^2)^{-1} \geq \sigma_{u0}^2 > 0$. Note that $\nabla_t^{\phi\phi}(f, \boldsymbol{\phi}) = \sigma_u^{-2} - 3\sigma_u^{-4} (y_t - f)^2$, so that $-\mathbb{E}[\nabla_t^{\phi\phi}(\omega_0, \boldsymbol{\phi}_0)] = 2\sigma_{u0}^{-2} \in (0, \infty)$. Fur-

thermore, $\Omega_{ff} - \mathbf{\Omega}_{\phi f}^\top (-\mathbb{E}[\nabla_t^{\phi\phi}(\omega_0, \phi_0)])^{-1} \mathbf{\Omega}_{\phi f} = \Omega_{ff} > 0$. Therefore, [Assumption 3\(v\)](#) is satisfied. Since $s_t(\omega_0, \phi_0) = u_t$ are independent, by [Corollary 2](#), (3.7) holds for QLR_T , namely $QLR_T \xrightarrow{d} \sup_{\beta \in \mathcal{B}} \left(\max \left\{ (1 - \beta^2)^{1/2} \sum_{j=0}^{\infty} \beta^j X_j, 0 \right\} \right)^2$.

C.2 t -GAS volatility

Consider the time-varying scale model of [Creal et al. \(2013\)](#), also examined in [Blasques et al. \(2022\)](#), $y_t = f_t^{1/2} u_t$, where u_t are i.i.d. t_ν random variables with $\nu > 0$ degrees of freedom. Set $\phi = \nu$. Then, $\nabla_t^f(f, \phi) = 2^{-1} f^{-2} \left(\frac{1 + \nu^{-1}}{1 + \nu^{-1} y_t^2 / f} y_t^2 - f \right)$. Taking $S_t(f, \phi) = 2f^2$, we obtain $s_t(f, \phi) = \frac{1 + \nu^{-1}}{1 + \nu^{-1} y_t^2 / f} y_t^2 - f$. Let $\mathcal{F} = \mathbb{R}^+$ and define $\mathcal{B} = [\beta_L, \beta_U] \subset (0, 1)$. Consider $\Theta = \{\boldsymbol{\vartheta} = (\alpha, \omega, \phi)^\top \in \mathbb{R}^3 : 0 \leq \alpha \leq \beta_L, 0 < \omega_L \leq \omega \leq \omega_U, 0 < \nu_L \leq \phi \leq \nu_U\}$. This specification guarantees the positivity of the filter. Furthermore, the filter is bounded below, with $f_t \geq \frac{\omega_L(1-\beta_U)}{1-\beta_U+\beta_L}$, $\forall t \in \mathbb{Z}$. We see that [Assumption 1](#) holds for any $\epsilon > 0$.

For [Assumption 2\(i\)](#), it suffices that $\{y_t, t \in \mathbb{Z}\}$ is an SE sequence. We then verify [Assumption 2\(ii\)](#) by checking the conditions in [Lemma A.1](#) and [Proposition 1](#). Since $(1+x)^{-2} \leq (1+2x)^{-1}$ for any $x \geq 0$, $\forall (\boldsymbol{\vartheta}, \beta) \in \Theta \times \mathcal{B}$, we obtain

$$\left| \beta + \alpha \frac{\partial}{\partial f} s_t(f, \phi) \right| \leq \beta + \alpha \nu \frac{y_t^2}{y_t^2 + 2\nu f} \leq \beta_U + \beta_L \nu_U \frac{y_t^2}{y_t^2 + c_0}, \quad (\text{C.3})$$

where $c_0 = 2\nu_L \omega_L (1 - \beta_U)(1 - \beta_U + \beta_L)^{-1} > 0$. By (C.3) and applying Jensen's inequality twice,

$$\begin{aligned} \mathbb{E} \left(\log \sup_{(\boldsymbol{\vartheta}, \beta) \in \Theta \times \mathcal{B}} \sup_{f \in \mathcal{F}} \left| \beta + \alpha \frac{\partial}{\partial f} s_t(f, \phi) \right| \right) &\leq \log \left(\beta_U + \beta_L \nu_U \mathbb{E} \left(\frac{y_t^2}{y_t^2 + c_0} \right) \right) \\ &\leq \log \left(\beta_U + \beta_L \nu_U \frac{\mathbb{E}(y_t^2)}{\mathbb{E}(y_t^2) + c_0} \right) < 0, \end{aligned} \quad (\text{C.4})$$

provided that the following condition holds:

$$\mathbb{E}(y_t^2) < \frac{2\nu_L \omega_L (1 - \beta_U)^2}{(1 - \beta_U + \beta_L)[\beta_L \nu_U - (1 - \beta_U)]}, \quad \beta_L \nu_U - (1 - \beta_U) > 0. \quad (\text{C.5})$$

Then Condition [IV3](#) holds with $r = 1$, which in turn ensures that Condition [IV2](#) also holds. Condition [IV1](#) is straightforward to verify and is therefore omitted. It follows that [Lemma A.1](#) holds. Note that Condition (C.5) is generally easy to verify. For practical implementation, the data may be scaled to ensure filter invertibility, after which the estimated filter can be rescaled to recover the original values. We now verify [Assumption 5](#) to establish [Proposition 1](#). Note

that $\sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta}_0^\epsilon \times \mathcal{B}} |s_t(f_t(\boldsymbol{\vartheta}, \beta), \phi)| \leq C \sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta}_0^\epsilon \times \mathcal{B}} |f_t(\boldsymbol{\vartheta}, \beta)|$. For $(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta}_0^\epsilon \times \mathcal{B}$, we have

$$\begin{aligned} f_{t+1}(\boldsymbol{\vartheta}, \beta) &= \omega(1 - \beta) + (\beta - \alpha)f_t(\boldsymbol{\vartheta}, \beta) + \alpha \frac{1 + \nu^{-1}}{1 + \nu^{-1}y_t^2/f_t(\boldsymbol{\vartheta}, \beta)} y_t^2 \\ &\leq \omega(1 - \beta) + (\beta + \alpha\nu)f_t(\boldsymbol{\vartheta}, \beta) \leq \omega_U(1 - \beta_L) + (\beta_U + \epsilon\nu_U)f_t(\boldsymbol{\vartheta}, \beta). \end{aligned}$$

Since [Assumption 1](#) imposes no restriction on $\epsilon > 0$, we can choose any $\epsilon \in (0, \nu_U^{-1}(1 - \beta_U))$ so that $\beta_U + \epsilon\nu_U < 1$, ensuring that $\sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta}_0^\epsilon \times \mathcal{B}} |f_t(\boldsymbol{\vartheta}, \beta)|$ is bounded. Compared to the condition $\sup_{\boldsymbol{\vartheta} \in \boldsymbol{\Theta}} (\beta + \nu\alpha) < 1$ imposed by [Blasques et al. \(2022, p. 331\)](#), we do not require such a strong restriction, as [Assumption 5](#) applies only to a local neighborhood $\boldsymbol{\Theta}_0^\epsilon$ of $\boldsymbol{\vartheta}_0$. Similarly, it is straightforward to show that $\sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta}_0^\epsilon \times \mathcal{B}} \left| \partial s_t(f, \phi) / \partial \phi \Big|_{f=f_t(\boldsymbol{\vartheta}, \beta)} \right|$ is also bounded. Hence, [Assumption 5\(i\)](#) holds for $\eta = \infty$. Using [\(C.3\)](#), we can set $\varrho_j(\boldsymbol{\vartheta}_0, \epsilon, \kappa) = (\beta_U + \epsilon\nu_U)^{j\kappa}$ in [Assumption 5\(ii\)](#). Thus, [Assumption 5\(ii\)](#) holds for any $\kappa \geq 1$ and some $\epsilon \in (0, \nu_U^{-1}(1 - \beta_U))$. Note that if we first compute the l/r partial derivatives of the mapping $(f, \phi) \mapsto s_t(f, \phi)$ and then substitute $f = f_t(\boldsymbol{\vartheta}, \beta)$, the resulting derivatives are bounded (locally) up to at least third order for all $(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta}_0^\epsilon \times \mathcal{B}$. By employing a mean value theorem (e.g., [Rudin, 1976](#), Theorem 9.19), [Assumption 5\(iii\)](#) is satisfied by setting $\zeta_\ell = 1$ and $C_t(\boldsymbol{\vartheta}_0, \epsilon) = C^\ddagger$ for some $C^\ddagger > 0$ for all $\ell = 0, 1, \dots, 5$ and $t \in \mathbb{Z}$. Moreover, the uniform boundedness condition in [Assumption 5\(iv\)](#) holds. Then, [Proposition 1](#) holds, with δ_1 and δ_2 able to take arbitrarily large values.

For [Assumption 3](#), and in particular [Assumptions 3\(iv\)–\(v\)](#), stronger conditions are required on the process $\{y_t, t \in \mathbb{Z}\}$. For simplicity, we next assume correct model specification under the null hypothesis and suppose that $y_t = \sqrt{\omega_0} u_t$, where $u_t \sim t_{\nu_0}$ are i.i.d. For [Assumption 3\(i\)](#), we have $\sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta} \times \mathcal{B}} |d_t(\boldsymbol{\vartheta}, \beta)| \xrightarrow{e.a.s.} 0$ by applying [Lemma A.1](#) and [Straumann and Mikosch \(2006, Lemma 2.1\)](#), together with a mean-value theorem. Moreover, $\mathbb{E}(\sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta} \times \mathcal{B}} |\ell_t(f_t(\boldsymbol{\vartheta}, \beta), \phi)|) < \infty$ holds if $\mathbb{E}(y_t^2) < \infty$ (which is implied by [\(C.5\)](#)) and if there exists some $\eta > 0$ such that $\mathbb{E}(\sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta} \times \mathcal{B}} |f_t(\boldsymbol{\vartheta}, \beta)|^\eta) < \infty$. Take $\eta = 1$. For $(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta} \times \mathcal{B}$, note that $|f_{t+1}(\boldsymbol{\vartheta}, \beta)| \leq |\psi_t(\omega, \boldsymbol{\vartheta}, \beta)| + |\psi_t(f_t(\boldsymbol{\vartheta}, \beta), \boldsymbol{\vartheta}, \beta) - \psi_t(\omega, \boldsymbol{\vartheta}, \beta)| \leq A_t(\boldsymbol{\vartheta}, \beta) + B_t |f_t(\boldsymbol{\vartheta}, \beta)|$, where $\psi_t(f, \boldsymbol{\vartheta}, \beta) = \omega(1 - \beta) + \beta f + \alpha \left(\frac{1 + \nu^{-1}}{1 + \nu^{-1}y_t^2/f} y_t^2 - f \right)$, $A_t(\boldsymbol{\vartheta}, \beta) = |\psi_t(\omega, \boldsymbol{\vartheta}, \beta)| + \sup_{f \in \mathcal{F}} |\partial \psi_t(f, \boldsymbol{\vartheta}, \beta) / \partial f| |\omega|$, and $B_t = \beta_U + \beta_L \nu_U y_t^2 / (y_t^2 + c_0)$ using [\(C.3\)](#). Since y_t is independent of $\mathcal{F}_{t-1}$ and $f_t(\boldsymbol{\vartheta}, \beta)$ is $\mathcal{F}_{t-1}$ -measurable ([Lemma A.1](#)), B_t and $f_t(\boldsymbol{\vartheta}, \beta)$ are independent, implying $\mathbb{E}(B_t \sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta} \times \mathcal{B}} |f_t(\boldsymbol{\vartheta}, \beta)|) = \mathbb{E}(B_t) \mathbb{E}(\sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta} \times \mathcal{B}} |f_t(\boldsymbol{\vartheta}, \beta)|)$. Under [\(C.5\)](#), we obtain $\mathbb{E}(\sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta} \times \mathcal{B}} A_t(\boldsymbol{\vartheta}, \beta)) < \infty$ and $\mathbb{E}(B_t) < 1$. Hence, $\mathbb{E}(\sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta} \times \mathcal{B}} |f_t(\boldsymbol{\vartheta}, \beta)|) < \infty$, and [Assumption 3\(i\)](#) follows. [Assumption 3\(ii\)](#) follows from [Lemma A.2](#). [Assumption 3\(iii\)](#) holds by applying [Proposition 1](#). By tedious linear algebra, together with the distributional assumption on

y_t and the results in [Lin and Lucas \(2025, Eq. \(E.6\)\)](#) and [Harvey \(2013, Proposition 39, p. 211\)](#), we obtain $\Omega_{ff} = \Sigma_{ff} = \nu_0/[2\omega_0^2(\nu_0 + 3)]$ and $\mathbf{\Omega}_{\phi f} = [\omega_0(\nu_0 + 3)(\nu_0 + 1)]^{-1} > 0$. Moreover, since $\mathbb{E}(s_t(\omega_0, \phi_0) | \mathcal{F}_{t-1}) = 0$, it follows that $\mathbb{E}(s_t(\omega_0, \phi_0)) = 0$, and hence $\mathbb{E}(h_{0,t}(\beta)) = 0$ for any $\beta \in \mathcal{B}$. All remaining conditions in Assumptions [3\(iv\)–\(v\)](#) and [Corollary 2](#) are satisfied. Therefore, QLR_T converges to the limiting distribution given in [\(3.7\)](#).

C.3 Poisson autoregression

Consider the Poisson time series model $y_t | \mathcal{F}_{t-1} \sim \text{Poisson}(f_t)$, where $f_t > 0$ for every $t \in \mathbb{Z}$. In this case, there is no static parameter ϕ , so all conditions involving ϕ and the corresponding components in [Theorem 1](#) can be disregarded, as noted below [Theorem 1](#). Moreover, in this example we allow for model mis-specification. That is, the statistician observes the count data $\{y_t, t = 1, \dots, T\}$, which need not follow a Poisson distribution in the dgp. We nevertheless require that $\{y_t, t \in \mathbb{Z}\}$ is an SE sequence with $\mathbb{E}|y_t|^4 < \infty$, $\mathbb{E}(y_t | \mathcal{F}_{t-1}) = \omega_0$, and $\mathbb{E}((y_t - \omega_0)^2 | \mathcal{F}_{t-1}) = \text{Var}(y_t)$. We obtain $f_{t+1} = \omega(1 - \beta) + \beta f_t + \alpha(y_t - f_t)$ since $\nabla_t^f(f, \phi) = (y_t - f)/f$ and by setting $S_t(f, \phi) = f$. This yields the well-known integer GARCH model of [Fokianos et al. \(2009\)](#). In what follows, we verify the conditions for [Theorem 1](#). As the verification closely parallels the Gaussian location example in [Section C.1](#), we focus only on the key differences.

Let $\mathcal{F} = \mathbb{R}^+$ and define $\mathcal{B} = [\beta_L, \beta_U] \subset (0, 1)$. Set $\Theta = \{\boldsymbol{\vartheta} = (\alpha, \omega)^\top \in \mathbb{R}^2 : 0 \leq \alpha \leq \beta_L, 0 < \omega_L \leq \omega \leq \omega_U\}$. The specifications of Θ and $\mathcal{B}$ ensure the positivity of f_t . Thus, [Assumption 1](#) holds. Furthermore, [Assumption 2](#) holds since [Lemma A.1](#) applies and [Proposition 1](#) holds with $\delta_1 = 4$ and $\delta_2 = 2$. It remains to verify [Assumption 3](#). Note that $\sup_{(\boldsymbol{\vartheta}, \beta) \in \Theta \times \mathcal{B}} |d_t(\boldsymbol{\vartheta}, \beta)| \leq D_{1,t} + D_{2,t}$, where $D_{1,t} := \sup_{(\boldsymbol{\vartheta}, \beta) \in \Theta \times \mathcal{B}} |\hat{f}_t(\boldsymbol{\vartheta}, \beta) - f_t(\boldsymbol{\vartheta}, \beta)|$ and $D_{2,t} := |y_t| \sup_{(\boldsymbol{\vartheta}, \beta) \in \Theta \times \mathcal{B}} |\log(\hat{f}_t(\boldsymbol{\vartheta}, \beta)) - \log(f_t(\boldsymbol{\vartheta}, \beta))|$. Since $\hat{f}_t(\boldsymbol{\vartheta}, \beta)$ and $f_t(\boldsymbol{\vartheta}, \beta)$ are both bounded below by $\omega_L(1 - \beta_U)$ for any $(\boldsymbol{\vartheta}, \beta) \in \Theta \times \mathcal{B}$, and using the inequality $|\log(x) - \log(y)| \leq |x - y|/(x \wedge y)$ for $x, y > 0$, we obtain $D_{2,t} \leq [\omega_L(1 - \beta_U)]^{-1} |y_t| D_{1,t}$. By [Lemma A.1](#) (which implies $D_{1,t} \xrightarrow{e.a.s.} 0$) and since $\mathbb{E}(\log^+ |y_t|) < \infty$, it follows from [Straumann and Mikosch \(2006, Lemma 2.1\)](#) that $D_{2,t} \xrightarrow{e.a.s.} 0$, and hence $\sup_{(\boldsymbol{\vartheta}, \beta) \in \Theta \times \mathcal{B}} |d_t(\boldsymbol{\vartheta}, \beta)| \xrightarrow{e.a.s.} 0$. Moreover, since $\log(y!) \leq y \log(y)$ for all $y \in \mathbb{Z}^+$, we have

$$\begin{aligned} \mathbb{E} \left(\sup_{(\boldsymbol{\vartheta}, \beta) \in \Theta \times \mathcal{B}} |\ell_t(f_t(\boldsymbol{\vartheta}, \beta), \phi)| \right) &\leq \mathbb{E} \left(|y_t| \sup_{(\boldsymbol{\vartheta}, \beta) \in \Theta \times \mathcal{B}} |\log(f_t(\boldsymbol{\vartheta}, \beta))| \right) \\ &\quad + \mathbb{E} \left(\sup_{(\boldsymbol{\vartheta}, \beta) \in \Theta \times \mathcal{B}} |f_t(\boldsymbol{\vartheta}, \beta)| \right) + \mathbb{E}(y_t \log(y_t)). \end{aligned}$$

We obtain $\mathbb{E}\left(\sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta} \times \mathcal{B}} |f_t(\boldsymbol{\vartheta}, \beta)|\right)^4 < \infty$ following the same steps as above (C.2). By the Cauchy–Schwarz inequality, it follows that $\mathbb{E}\left(\sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta} \times \mathcal{B}} |\ell_t(f_t(\boldsymbol{\vartheta}, \beta), \boldsymbol{\phi})|\right) < \infty$, and thus [Assumption 3\(i\)](#) holds. Consider [Assumption 3\(ii\)](#) next. By applying Jensen’s inequality to the function $f \mapsto \omega_0 \log(f) - f$ for $f > 0$ and $\omega_0 > 0$, we obtain

$$\begin{aligned} \mathbb{E}\left(\ell_t(f_t(\boldsymbol{\vartheta}, \beta), \boldsymbol{\phi})\right) &= \mathbb{E}\left(\omega_0 \log(f_t(\boldsymbol{\vartheta}, \beta)) - f_t(\boldsymbol{\vartheta}, \beta)\right) - \mathbb{E}(\log(y_t!)) \\ &\leq \left(\omega_0 \log\left(\mathbb{E}(f_t(\boldsymbol{\vartheta}, \beta))\right) - \mathbb{E}(f_t(\boldsymbol{\vartheta}, \beta))\right) - \mathbb{E}(\log(y_t!)), \end{aligned}$$

where equality holds if and only if $f_t(\boldsymbol{\vartheta}, \beta) = \mathbb{E}(f_t(\boldsymbol{\vartheta}, \beta)) > 0$ a.s., that is, $f_t(\boldsymbol{\vartheta}, \beta)$ is a.s. constant. Since $s_t(f, \boldsymbol{\phi}) = y_t - f_t$ is random, one must have $\alpha = 0 = \alpha_0$, which implies $f_t(\boldsymbol{\vartheta}, \beta) = \omega$ for any $\beta \in \mathcal{B}$ a.s. Hence, $\mathbb{E}(\ell_t(f_t(\boldsymbol{\vartheta}, \beta), \boldsymbol{\phi})) \leq \omega_0 \log(\omega) - \omega - \mathbb{E}(\log(y_t!)) \leq \omega_0 \log(\omega_0) - \omega_0 - \mathbb{E}(\log(y_t!))$, with equality if and only if $\omega = \omega_0$. [Assumption 3\(ii\)](#) is thereby verified. Similarly, [Assumption 3\(iii\)](#) is satisfied, and [Assumption 3\(iv\)](#) holds with $\Sigma_{ff} = \text{Var}(y_t)/\omega_0^2 < \infty$, $\Omega_{ff} = -\mathbb{E}(\nabla_t^{ff}(\omega_0, \boldsymbol{\phi}_0)) = \omega_0^{-1} \in (0, \infty)$. Finally, we have $\inf_{\beta \in \mathcal{B}} \text{Var}\left(\sum_{j=0}^{\infty} \beta^j s_t(\omega_0, \boldsymbol{\phi}_0)\right) = (1 - \beta_L^2)^{-1} \text{Var}(y_t) > 0$. Therefore, [Assumption 3\(v\)](#) is fulfilled. By [Corollary 1](#), (3.7) holds for $\widetilde{QLR}_T$ provided that $\hat{\kappa}_{G,n} = \text{Var}(y_t)/\omega_0 + o_{\mathbb{P}}(1)$. If, in addition, $\text{Var}(y_t) = \omega_0$, then QLR_T converges to the same limiting distribution as $\widetilde{QLR}_T$ by [Corollary 2](#).

C.4 Spatial dynamic spillovers

Consider the model of [Blasques et al. \(2016\)](#): $\mathbf{y}_t = \tilde{\rho}(f_t) \mathbf{W} \mathbf{y}_t + \mathbf{x}_t \boldsymbol{\gamma} + \mathbf{e}_t$, where $\mathbf{W}$ is an $n \times n$ matrix of exogenous, nonrandom spatial weights, $\mathbf{x}_t$ is a $n \times k$ matrix of regressors, $\boldsymbol{\gamma}$ is an $k \times 1$ vector of static parameters, and $\mathbf{e}_t$ is an $n \times 1$ disturbance vector. Here, both n and k are finite. The link function is specified as $\tilde{\rho}(\cdot) := \rho \tanh(\cdot)$, where $\rho \in (0, 1)$ is a user-specified parameter. Note that $\tilde{\rho}(f) \in (-\rho, \rho)$ for all $f \in \mathcal{F}$, and that its derivatives of every order are bounded. Furthermore, to ensure the invertibility of $\mathbf{I}_n - \tilde{\rho}(f) \mathbf{W}$, we require the spectral radius of $\mathbf{W}$ to be strictly smaller than 1. Suppose the statistician specifies that $\mathbf{e}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{V})$ are i.i.d., where $\mathbf{V} = \text{diag}(\sigma_1^2, \dots, \sigma_n^2)$ is positive definite. Let $\boldsymbol{\phi} = (\boldsymbol{\gamma}^\top, \sigma_1, \dots, \sigma_n)^\top$. We have $\ell_t(f, \boldsymbol{\phi}) = -2^{-1}n \log(2\pi) + \log \det(\mathbf{I}_n - \tilde{\rho}(f) \mathbf{W}) - 2^{-1} \log \det(\mathbf{V}) - 2^{-1}(\mathbf{y}_t - \tilde{\rho}(f) \mathbf{W} \mathbf{y}_t - \mathbf{x}_t \boldsymbol{\gamma})^\top \mathbf{V}^{-1}(\mathbf{y}_t - \tilde{\rho}(f) \mathbf{W} \mathbf{y}_t - \mathbf{x}_t \boldsymbol{\gamma})$. Take $S_t(f, \boldsymbol{\phi}) = 1$. Then the forcing variable $s_t(f, \boldsymbol{\phi})$ in (2.1) is given by $s_t(f, \boldsymbol{\phi}) = \nabla_t^f(f, \boldsymbol{\phi}) = \left(\mathbf{y}_t^\top \mathbf{W}^\top \mathbf{V}^{-1}[\mathbf{y}_t - \tilde{\rho}(f) \mathbf{W} \mathbf{y}_t - \mathbf{x}_t \boldsymbol{\gamma}] - \text{tr}(\mathbf{Z}(f) \mathbf{W})\right) \dot{\tilde{\rho}}(f)$, where $\mathbf{Z}(f) := (\mathbf{I}_n - \tilde{\rho}(f) \mathbf{W})^{-1}$ and $\dot{\tilde{\rho}}(f) = \rho(1 - \tanh^2(f))$.

To study the null distribution of QLR_T , we assume that $\mathbf{y}_t$ is generated as $\mathbf{y}_t = \tilde{\rho}(\omega_0) \mathbf{W} \mathbf{y}_t + \mathbf{x}_t \boldsymbol{\gamma}_0 + \boldsymbol{\varepsilon}_t$, where the error term $\boldsymbol{\varepsilon}_t := (\varepsilon_{1,t}, \dots, \varepsilon_{n,t})^\top$ has zero mean, and $\mathbb{E}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top \mid \mathbf{x}_t, \mathcal{F}_{t-1}) = \mathbf{V}_0$,

with $\mathbf{V}_0 = \text{diag}(\sigma_{0,1}^2, \dots, \sigma_{0,n}^2)$ diagonal and positive definite. However, we do not require $\boldsymbol{\varepsilon}_t$ to follow the same distribution as $\mathbf{e}_t$. We require that $\{(\mathbf{y}_t, \mathbf{x}_t), t \in \mathbb{Z}\}$ is SE with $\mathbb{E}(\|\boldsymbol{\varepsilon}_t\|^4 + \|\mathbf{x}_t\|^4) < \infty$, $\mathbb{E}(\mathbf{x}_t^\top \mathbf{x}_t)$ is positive definite, and $\mathbb{E}[\mathbf{y}_t^\top \mathbf{W}^\top \mathbf{V}_0^{-1} \boldsymbol{\varepsilon}_t - \text{tr}(\mathbf{Z}(\omega_0) \mathbf{W})]^2 > 0$. Moreover, $\mathbb{E}(\boldsymbol{\varepsilon}_t \mid \mathbf{x}_t, \mathcal{F}_{t-1}) = \mathbf{0}$ a.s., and $\mathbb{E}(\boldsymbol{\varepsilon}_t^\top \otimes (\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top) \mid \mathbf{x}_t, \mathcal{F}_{t-1})$, $\mathbb{E}((\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top) \otimes (\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top) \mid \mathbf{x}_t, \mathcal{F}_{t-1})$, $\mathbb{E}(\mathbf{x}_t \mid \mathcal{F}_{t-1})$, and $\mathbb{E}(\mathbf{x}_t^\top \otimes \mathbf{x}_t \mid \mathcal{F}_{t-1})$ are all nonrandom and bounded a.s., where $\otimes$ denotes the Kronecker product. We impose a technical assumption that there exist $1 \leq k < i \leq n$ such that $\sigma_{0,i}^2 [\mathbf{W} \mathbf{Z}(\omega_0)]_{ki} + \sigma_{0,k}^2 [\mathbf{W} \mathbf{Z}(\omega_0)]_{ik} \neq 0$, where $[\mathbf{W} \mathbf{Z}(\omega_0)]_{ki}$ denotes the (k, i) th element of $\mathbf{W} \mathbf{Z}(\omega_0)$, and similarly for $[\mathbf{W} \mathbf{Z}(\omega_0)]_{ik}$. Set $\mathcal{F} = \mathbb{R}$, $\mathcal{B} = [\beta_L, \beta_U] \subset (-1, 1)$, and $\boldsymbol{\Theta} = \{\boldsymbol{\vartheta} = (\alpha, \omega, \boldsymbol{\gamma}^\top, \sigma_1, \dots, \sigma_n)^\top \in \mathbb{R}^{k+n+2} : \alpha_L \leq \alpha \leq \alpha_U, \omega_L \leq \omega \leq \omega_U, \boldsymbol{\gamma}_L \preceq \boldsymbol{\gamma} \preceq \boldsymbol{\gamma}_U, 0 < \sigma_L \leq \sigma_i \leq \sigma_U, i = 1, \dots, n\}$, where $\alpha_L < 0$. Verifying [Assumption 2\(ii\)](#) requires tedious linear algebra. For illustration, we simply assume these conditions are satisfied, with $\mathbb{E}\left(\sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta}_\epsilon \times \mathcal{B}} \|\mathbf{f}_t^{(1)}(\boldsymbol{\vartheta}, \beta)\|^4\right) < \infty$ and $\mathbb{E}\left(\sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta}_\epsilon \times \mathcal{B}} \|\mathbf{f}_t^{(2)}(\boldsymbol{\vartheta}, \beta)\|^2\right) < \infty$ for some $\epsilon > 0$, and refer interested readers to [Blasques et al. \(2016\)](#) for a more detailed theoretical analysis. We also require that the identification condition in [Assumption 3\(ii\)](#) holds.

It follows immediately that [Assumption 1](#) holds for any $\epsilon > 0$. [Assumption 2\(i\)](#) holds by assumption. We now verify [Assumption 3](#). Given the moment conditions above and the fact that $n, k < \infty$, it is straightforward to show that $\mathbb{E}\left(\log^+ \sup_{(f, \boldsymbol{\vartheta}) \in \mathcal{F} \times \boldsymbol{\Theta}} |\nabla_t^f(f, \boldsymbol{\vartheta})|\right) < \infty$. Consequently, $\sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta} \times \mathcal{B}} |d_t(\boldsymbol{\vartheta}, \beta)| \xrightarrow{e.a.s.} 0$ follows from [Assumption 2\(ii\)](#) and [Straumann and Mikosch \(2006, Lemma 2.1\)](#). Moreover, since $n < \infty$ and the spectral radius of the nonrandom matrix $\mathbf{W}$ is strictly smaller than ρ^{-1} , we have $\mathbb{E}\left(\sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta} \times \mathcal{B}} |\log \det(\mathbf{I}_n - h(f_t(\boldsymbol{\vartheta}, \beta)) \mathbf{W})|\right) < \infty$. Using $\mathbb{E}\|\mathbf{y}_t\|^2 + \mathbb{E}\|\mathbf{x}_t\|^2 < \infty$, it follows that $\mathbb{E}\left(\sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta} \times \mathcal{B}} |\ell_t(f_t(\boldsymbol{\vartheta}, \beta), \boldsymbol{\varphi})|\right) < \infty$. We obtain [Assumption 3\(i\)](#). To proceed, let $\mathbf{e}_t(f, \boldsymbol{\gamma}) = (e_{1,t}(f, \boldsymbol{\gamma}), \dots, e_{n,t}(f, \boldsymbol{\gamma}))^\top = \mathbf{y}_t - \tilde{\rho}(f) \mathbf{W} \mathbf{y}_t - \mathbf{x}_t \boldsymbol{\gamma}$. Since $\mathbb{E}(\|\boldsymbol{\varepsilon}_t\|^4 + \|\mathbf{x}_t\|^4) < \infty$, it follows that $\mathbb{E}(\|\mathbf{y}_t\|^4) \leq 8\|\mathbf{Z}(\omega_0)\|^4 \mathbb{E}(\|\mathbf{x}_t\|^4 \|\boldsymbol{\gamma}_0\|^4 + \|\boldsymbol{\varepsilon}_t\|^4) < \infty$, and $\mathbb{E}\left(\sup_{(\boldsymbol{\vartheta}, \beta) \in \boldsymbol{\Theta}_\epsilon \times \mathcal{B}} \|\mathbf{e}_t(f_t(\boldsymbol{\vartheta}, \beta), \boldsymbol{\gamma})\|^4\right) < \infty$ for some $\epsilon > 0$. By repeatedly applying the Cauchy-Schwarz inequality, together with the above moment conditions and the quantities computed below, one can establish [Assumption 3\(iii\)](#).

Now we turn to Assumptions [3\(iv\)](#)–[\(v\)](#). The straightforward steps are omitted. Let $\mathbb{E}_{t-1}(\cdot) = \mathbb{E}(\cdot \mid \mathbf{x}_t, \mathcal{F}_{t-1})$. Observe that $\mathbb{E}_{t-1}(\nabla_t^f(\omega_0, \boldsymbol{\phi}_0)) = \left[\text{tr}\left(\mathbf{Z}(\omega_0)^\top \mathbf{W}^\top \mathbf{V}_0^{-1} \mathbb{E}_{t-1}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^\top)\right) - \text{tr}(\mathbf{Z}(\omega_0) \mathbf{W}) \right] \dot{\rho}(\omega_0) = 0$. This immediately implies that $\mathbb{E}(h_{0,t}(\beta)) = 0$ for any $\beta \in \mathcal{B}$. Moreover, since $\nabla_t^{ff}(f, \boldsymbol{\phi}) = -\left[\mathbf{y}_t^\top \mathbf{W}^\top \mathbf{V}^{-1} \mathbf{W} \mathbf{y}_t + \text{tr}\left((\mathbf{Z}(f) \mathbf{W})^2\right)\right] (\dot{\rho}(f))^2 + \left[\mathbf{y}_t^\top \mathbf{W}^\top \mathbf{V}^{-1} \mathbf{e}_t(f, \boldsymbol{\gamma}) - \text{tr}(\mathbf{Z}(f) \mathbf{W})\right] \ddot{\rho}(f)$, we have $\Omega_{ff} = \left[\mathbb{E}_{t-1}\left(\mathbf{y}_t^\top \mathbf{W}^\top \mathbf{V}_0^{-1} \mathbf{W} \mathbf{y}_t\right) + \text{tr}\left((\mathbf{Z}(\omega_0) \mathbf{W})^2\right)\right] (\dot{\rho}(\omega_0))^2 \in (0, \infty)$ which is nonrandom by assumption. One also has $\Sigma_{ff} = \mathbb{E}_{t-1}([\nabla_t^f(\omega_0, \boldsymbol{\phi}_0)]^2) \in (0, \infty)$ is nonrandom. Since $\partial \ell_t(f, \boldsymbol{\phi}) / \partial \boldsymbol{\gamma} = \mathbf{x}_t^\top \mathbf{V}^{-1} \mathbf{e}_t(f, \boldsymbol{\gamma})$ and $\partial \ell_t(f, \boldsymbol{\phi}) / \partial \sigma_i = \sigma_i^{-2} [\sigma_i^{-2} e_{i,t}^2(f, \boldsymbol{\gamma}) - 1]$, where $i = 1, \dots, n$, we have $\mathbb{E}_{t-1}(\nabla_t^\phi(\omega_0, \boldsymbol{\phi}_0)) = \mathbf{0}$, and $\boldsymbol{\Sigma}_{\phi f} = \mathbb{E}_{t-1}(\nabla_t^\phi(\omega_0, \boldsymbol{\phi}_0) \nabla_t^f(\omega_0, \boldsymbol{\phi}_0))$

is nonrandom. Note that $\partial \nabla_t^f(f, \phi) / \partial \gamma = -\mathbf{x}_t^\top \mathbf{V}^{-1} \mathbf{W} \mathbf{y}_t \dot{\rho}(f) = -\mathbf{x}_t^\top \mathbf{V}^{-1} \mathbf{W} \mathbf{Z}(\omega_0) \boldsymbol{\varepsilon}_t \dot{\rho}(f) - \mathbf{x}_t^\top \mathbf{V}^{-1} \mathbf{W} \mathbf{Z}(\omega_0) \mathbf{x}_t \gamma_0 \dot{\rho}(f)$, $\partial \nabla_t^f(f, \phi) / \partial \sigma_i = -2\mathbf{e}_t(f, \gamma)^\top \mathbf{V}^{-1} \mathcal{P}_n^{(i)} \mathbf{V}^{-1} \mathbf{W} \mathbf{Z}(\omega_0) \mathbf{x}_t \gamma_0 \sigma_i \dot{\rho}(f) - 2\mathbf{e}_t(f, \gamma)^\top \mathbf{V}^{-1} \mathcal{P}_n^{(i)} \mathbf{V}^{-1} \mathbf{W} \mathbf{Z}(\omega_0) \boldsymbol{\varepsilon}_t \sigma_i \dot{\rho}(f)$ for $i = 1, \dots, n$. Here, $\mathcal{P}_K^{(j)} \in \mathbb{R}^{K \times K}$ denotes a matrix with zeros everywhere except for a single 1 at the j th diagonal position for $j = 1, \dots, K$, so that $\mathcal{P}_K^{(j)} \mathbf{A}$ acts as a row selector, zeroing out all but the j th row of a $K \times K$ matrix $\mathbf{A}$. Then,

$$\boldsymbol{\Omega}_{\phi f} = \mathbb{E}_{t-1}(\nabla_t^{\phi f}(\omega_0, \phi_0)) = -\dot{\rho}(\omega_0) \begin{pmatrix} \mathbb{E}_{t-1}(\mathbf{x}_t^\top \mathbf{V}_0^{-1} \mathbf{W} \mathbf{Z}(\omega_0) \mathbf{x}_t) \gamma_0 \\ 2 \operatorname{tr}(\mathcal{P}_n^{(1)} \mathbf{W} \mathbf{Z}(\omega_0)) \sigma_{0,1}^{-1} \\ \vdots \\ 2 \operatorname{tr}(\mathcal{P}_n^{(n)} \mathbf{W} \mathbf{Z}(\omega_0)) \sigma_{0,n}^{-1} \end{pmatrix},$$

which is nonrandom. Thus, [Assumption 3\(iv\)](#) holds. Note that $\inf_{\beta \in \mathcal{B}} \mathbb{V} \operatorname{ar}(h_{0,t}(\beta)) \geq \Sigma_{ff} > 0$, and $-\mathbb{E}(\nabla_t^{\phi \phi}(\omega_0, \phi_0)) = \operatorname{diag}(\mathbb{E}(\mathbf{x}_t^\top \mathbf{V}_0^{-1} \mathbf{x}_t), 2\mathbf{V}_0^{-1})$. Then, $0 < \lambda_{\min}(-\mathbb{E}[\nabla_t^{\phi \phi}(\omega_0, \phi_0)]) \leq \lambda_{\max}(-\mathbb{E}[\nabla_t^{\phi \phi}(\omega_0, \phi_0)]) < \infty$ because $\mathbb{E}(\mathbf{x}_t^\top \mathbf{x}_t)$ is positive definite and $\mathbb{E}\|\mathbf{x}_t\|^2 < \infty$. We further have $\Omega_{ff} - \boldsymbol{\Omega}_{\phi f}^\top (-\mathbb{E}[\nabla_t^{\phi \phi}(\omega_0, \phi_0)])^{-1} \boldsymbol{\Omega}_{\phi f} = (\gamma_0^\top \mathbf{R}_1 \gamma_0 + R_2) (\dot{\rho}(\omega_0))^2$, where

$$\begin{aligned} \mathbf{R}_1 &:= \mathbb{E}_{t-1}(\mathbf{x}_t^\top \mathbf{Z}(\omega_0)^\top \mathbf{W}^\top \mathbf{V}_0^{-1} \mathbf{W} \mathbf{Z}(\omega_0) \mathbf{x}_t) \\ &\quad - \mathbb{E}_{t-1}(\mathbf{x}_t^\top \mathbf{Z}(\omega_0)^\top \mathbf{W}^\top \mathbf{V}_0^{-1} \mathbf{x}_t) \left[\mathbb{E}(\mathbf{x}_t^\top \mathbf{V}_0^{-1} \mathbf{x}_t) \right]^{-1} \mathbb{E}_{t-1}(\mathbf{x}_t^\top \mathbf{V}_0^{-1} \mathbf{W} \mathbf{Z}(\omega_0) \mathbf{x}_t), \\ R_2 &:= \operatorname{tr}(\mathbf{Z}(\omega_0)^\top \mathbf{W}^\top \mathbf{V}_0^{-1} \mathbf{W} \mathbf{Z}(\omega_0) \mathbf{V}_0) + \operatorname{tr}((\mathbf{Z}(\omega_0) \mathbf{W})^2) - 2 \sum_{i=1}^n \left[\operatorname{tr}(\mathcal{P}_n^{(i)} \mathbf{W} \mathbf{Z}(\omega_0)) \right]^2. \end{aligned}$$

It remains to show that $\Omega_{ff} - \boldsymbol{\Omega}_{\phi f}^\top (-\mathbb{E}[\nabla_t^{\phi \phi}(\omega_0, \phi_0)])^{-1} \boldsymbol{\Omega}_{\phi f} > 0$. Let $\mathbf{U}_t := \mathbf{V}_0^{-1/2} \mathbf{x}_t$ and $\mathbf{X}_t := \mathbf{V}_0^{-1/2} \mathbf{W} \mathbf{Z}(\omega_0) \mathbf{x}_t$. Then $\mathbf{R}_1 = \mathbb{E}_{t-1}(\mathbf{X}_t^\top \mathbf{X}_t) - \mathbb{E}_{t-1}(\mathbf{X}_t^\top \mathbf{U}_t) [\mathbb{E}(\mathbf{U}_t^\top \mathbf{U}_t)]^{-1} \mathbb{E}_{t-1}(\mathbf{U}_t^\top \mathbf{X}_t)$, which is positive semidefinite because $\begin{pmatrix} \mathbb{E}(\mathbf{U}_t^\top \mathbf{U}_t) & \mathbb{E}_{t-1}(\mathbf{U}_t^\top \mathbf{X}_t) \\ \mathbb{E}_{t-1}(\mathbf{X}_t^\top \mathbf{U}_t) & \mathbb{E}_{t-1}(\mathbf{X}_t^\top \mathbf{X}_t) \end{pmatrix} =_{a.s.} \mathbb{E}_{t-1} \begin{pmatrix} \mathbf{U}_t^\top \mathbf{U}_t & \mathbf{U}_t^\top \mathbf{X}_t \\ \mathbf{X}_t^\top \mathbf{U}_t & \mathbf{X}_t^\top \mathbf{X}_t \end{pmatrix}$ is positive semidefinite; see [Abadir and Magnus \(2005, Exercise 12.33, p. 341\)](#). For R_2 , let $\mathbf{Q} := \mathbf{W} \mathbf{Z}(\omega_0) = (q_{ij}, 1 \leq i, j \leq n)$. Since $\operatorname{tr}(\mathbf{Z}(\omega_0) \mathbf{W} \mathbf{Z}(\omega_0) \mathbf{W}) = \operatorname{tr}(\mathbf{W} \mathbf{Z}(\omega_0) \mathbf{W} \mathbf{Z}(\omega_0))$, we have

$$\begin{aligned} R_2 &= \operatorname{tr}(\mathbf{Q}^\top \mathbf{V}_0^{-1} \mathbf{Q} \mathbf{V}_0) + \operatorname{tr}(\mathbf{Q}^2) - 2 \sum_{i=1}^n \left[\operatorname{tr}(\mathcal{P}_n^{(i)} \mathbf{Q}) \right]^2 \\ &= \sum_{1 \leq k \neq i \leq n} \left[\left(\frac{\sigma_{0,i}}{\sigma_{0,k}} q_{ki} \right)^2 + q_{ki} q_{ik} \right] \\ &= \sum_{1 \leq k < i \leq n} \left(\frac{\sigma_{0,i}^2}{\sigma_{0,k}^2} q_{ki}^2 + 2q_{ki} q_{ik} + \frac{\sigma_{0,k}^2}{\sigma_{0,i}^2} q_{ik}^2 \right) = \sum_{1 \leq k < i \leq n} \left(\frac{\sigma_{0,i}}{\sigma_{0,k}} q_{ki} + \frac{\sigma_{0,k}}{\sigma_{0,i}} q_{ik} \right)^2 \geq 0. \end{aligned}$$

If there exist $1 \leq k < i \leq n$ such that $\sigma_{0,i}^2 q_{ki} + \sigma_{0,k}^2 q_{ik} \neq 0$, then $R_2 > 0$, and consequently $\Omega_{ff} - \boldsymbol{\Omega}_{\phi f}^\top (-\mathbb{E}[\nabla_t^{\phi \phi}(\omega_0, \phi_0)])^{-1} \boldsymbol{\Omega}_{\phi f} > 0$. This completes the verification of [Assumption 3\(v\)](#).

Since Condition (i) holds as established above, if Condition (ii) is also satisfied, then Corollary 1 yields $\widetilde{QLR}_T \xrightarrow{d} \sup_{\beta \in \mathcal{B}} \left((1 - \beta^2)^{1/2} \sum_{j=0}^{\infty} \beta^j X_j \right)^2$. Under correct model specification, Corollary 2 further shows that QLR_T converges to the same limiting distribution.

D Parameter estimates for the empirical example

Table D.4: Parameter estimates obtained by joint optimization of $(\boldsymbol{\vartheta}, \beta)$ under the assumption of Gaussian errors for the empirical example in Section 6, with the rows labeled “restricted” representing the estimates under the condition $\alpha = 0$.

	α	β	ω	intercept	Δ VSTOXX returns	Δ Eonia 3M changes	stock index index	yield spread spread
restricted	0	0.5254	0.1821	−0.0065	−0.5153	3.1231	−6.8573	−0.0184
unrestricted	$-6.8555 \cdot 10^{-4}$	0.9949	0.1374	−0.0073	−0.4197	3.1761	−6.7676	−0.0404

	σ						
	Germany	France	Ireland	Italy	Netherlands	Portugal	Spain
restricted	0.9009	1.5150	2.3082	8.3304	0.8797	12.0016	5.3201
unrestricted	0.8663	1.4646	2.3039	8.3245	0.8669	11.9934	5.3404

Table D.5: Parameter estimates obtained by joint optimization of $(\boldsymbol{\vartheta}, \beta)$ under the assumption of Student’s t errors for the empirical example in Section 6, with the rows labeled “restricted” representing the estimates under the condition $\alpha = 0$.

	α	β	ω	ν	intercept	Δ VSTOXX returns	Δ Eonia 3M changes	stock index index	yield spread spread
restricted	0	0.5028	0.1475	1.1701	−0.0124	0.1123	1.1637	−1.5749	−0.2239
unrestricted	0.0075	0.6294	0.1579	1.1768	−0.0120	0.0868	1.3502	−1.4848	−0.2351

	σ						
	Germany	France	Ireland	Italy	Netherlands	Portugal	Spain
restricted	0.3590	3.2060	0.5203	0.7315	3.2474	0.2927	1.7827
unrestricted	0.3628	3.2130	0.5186	0.7310	3.2815	0.2903	1.7825

References in Online Appendix

- Abadir, K. M. and J. R. Magnus (2005). *Matrix Algebra*. Cambridge: Cambridge University Press.
- Blasques, F., S. J. Koopman, A. Lucas, and J. Schaumburg (2016). Spillover dynamics for systemic risk measurement using spatial financial time series models. *Journal of Econometrics* 195(2), 211–223.
- Blasques, F., J. van Brummelen, S. J. Koopman, and A. Lucas (2022). Maximum likelihood estimation for score-driven models. *Journal of Econometrics* 227(2), 325–346.
- Bougerol, P. (1993). Kalman filtering with random coefficients and contractions. *SIAM Journal on Control and Optimization* 31(4), 942–959.
- Creal, D., S. J. Koopman, and A. Lucas (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics* 28(5), 777–795.
- Fokianos, K., A. Rahbek, and D. Tjøstheim (2009). Poisson autoregression. *Journal of the American Statistical Association* 104(488), 1430–1439.
- Harvey, A. C. (2013). *Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series*. Econometric Series Monographs. Cambridge: Cambridge University Press.
- Lin, Y. and A. Lucas (2025). Functional location-scale models with robust observation-driven dynamics. Technical Report TI 2025-027/III, Tinbergen Institute Discussion Paper.
- Rudin, W. (1976). *Principles of Mathematical Analysis* (3rd ed.). New York: McGraw-Hill.
- Straumann, D. and T. Mikosch (2006). Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: A stochastic recurrence equations approach. *Annals of Statistics* 34(5), 2449–2495.
- White, H. (1996). *Estimation, Inference and Specification Analysis*. Cambridge University Press.