TI 2026-060/III
Tinbergen Institute Discussion Paper

Joint Eigenvector and Eigenvalue Dynamics with an Application to Time-Varying Covariance Matrices

Justus Holman[1]
Yicong Lin[1]
Andre Lucas[1]
Anne Opschoor[1]

1 Vrije Universiteit Amsterdam and Tinbergen Institute

Joint Eigenvector and Eigenvalue Dynamics with an Application to Time-Varying Covariance Matrices

Justus Holman* Yicong Lin* Andre Lucas* Anne Opschoor*

August, 2026

Abstract

We introduce the Dynamic Spectral Rotation (DSR) model, allowing for dynamics in both eigenvalues and eigenvectors of time-varying conditional covariance matrices. The construction preserves orthonormality of the entire eigenvector matrix at every point in time. We study the model's asymptotic properties and establish unique identification of all static parameters governing the joint dynamics of eigenvalues and eigenvectors. Notably, the parameters that determine the dynamic rotation angles remain uniquely identified under mild conditions even when the rotation angles are allowed to evolve over ranges far beyond intervals of length π. In an empirical application to US equity returns, we show that allowing the leading eigendirection of the covariance matrix to vary over time significantly improves the predicted portfolio covariance structure compared to a model in which all eigendirections are held fixed.

Keywords: dynamic covariance matrices; dynamic eigenvectors; Givens rotation; score-driven dynamics; volatility modeling.

*Vrije Universiteit Amsterdam and Tinbergen Institute, the Netherlands. Holman and Opschoor thank the Dutch National Science Foundation (NWO) for financial support under grant VI.VIDI.201.079.

1 Introduction

Accurately modeling dynamic conditional covariance structures has been one of the dominant research areas in econometrics over the past decades, leading to a range of well-known models such as the BEKK model (Engle and Kroner, 1995), the Constant Conditional Correlations model (Bollerslev, 1990), the Dynamic Conditional Correlation model (Engle, 2002), and many others.

One of the major challenges in this line of literature is to keep the model flexible enough for the empirical data at hand, while at the same time limiting the number of parameters. A recent promising direction in this respect is the λ-GARCH model introduced by Hetland et al. (2023). They propose a highly parsimonious way to capture conditional covariance matrix dynamics via an eigenvalue-eigenvector decomposition of the covariance matrix, where the eigenvalues have score-driven dynamics (Creal et al., 2013; Harvey, 2013). The intuition is that eigenvalues capture the most prominent axes of variation for a dynamic covariance matrix. By allowing the dynamics of one eigenvalue to spill over into those of the others, the model provides more flexibility compared to the specifications of Van der Weide (2002) and Bollerslev (1990). Gubbels and Lucas (2026) extend the model to include a larger number of assets and combine this with the shrinkage techniques of Ledoit and Wolf (2022) for correcting the eigenvalue spectrum.

While the eigenvalues in the model of Hetland et al. (2023) are allowed to vary over time, the eigenvectors are taken as constant. Though convenient, such an assumption may be too restrictive for typical empirical data. For instance, it is likely that allowing for flexibility in the direction of the first or first few eigenvectors over time may substantially increase the model's fit to the data, as these directions correspond to the largest eigenvalues and, thus, the largest sources of variation. The major challenge in introducing dynamics into the eigenvectors is that the eigenvectors of a covariance matrix need to constitute an orthogonal basis at all time points.

Therefore, dynamically re-orienting one eigenvector, such as the first one, requires the remaining eigenvectors to adjust accordingly to preserve the orthonormality of the full matrix of eigenvectors. This is precisely the problem we address in the current paper.

The new model we propose in this paper allows for both dynamic eigenvalues and eigenvectors, where the eigenvector matrix remains orthonormal by construction at every time point. In particular, we introduce time variation in the eigenvectors by dynamically rotating a set of target eigenvectors via a product of Givens rotation matrices. These rotation matrices are driven by time-varying parameters that we endow with score-driven dynamics as in Creal et al. (2013) and Harvey (2013). This updating mechanism for the eigenvectors preserves the necessary orthonormality of the updated eigenvectors by rotating the entire basis spanned by them. In this paper, we focus on the dynamics of the first eigenvector, as this one corresponds to the direction in which the data shows the most variation. It is here that we expect to realize most gains with the new model specification. For this version of the model, we develop the full theoretical framework.

The resulting Dynamic Spectral Rotation (DSR) model captures the most important dynamics in the covariance matrix, or, if variances have been filtered out first, the correlation matrix. In our theoretical analysis of the new model, we establish conditions under which the model is identified, and the maximum likelihood estimator is consistent and asymptotically normal. Surprisingly, we find that the paths of the dynamic rotation angles remain identified even when they are allowed to evolve beyond a commonly imposed identifying range of length π such as $(0, \pi]$. In fact, we establish that the *intercepts* in the transition equation for the rotation angles have to satisfy such a restriction, but that the angles themselves can move freely beyond this range. The surprising element is that at any specific moment the angles can indeed be rotated arbitrarily by a multiple of π. The score-driven dynamics, however, heavily restrict such rotations over time. Our new dynamic rather than static identification strategy leverages this phenomenon and is used to establish the new result. Simulations further illustrate that the paths of the angles can indeed be recovered over considerable ranges if the model is applied to data.

In our empirical application, we study 10 series of daily stock returns for 10 US financials from 2002 to 2023. Empirically, it turns out to be important to allow the first eigenvector to vary over time. Across a range of statistical and economic performance criteria, the new model improves the out-of-sample performance of the λ-GARCH benchmark of Hetland et al. (2023), which allows only for dynamic eigenvalues. The changes in the eigendirections in the DSR model closely align with major economic events, changing the dominant systematic risk direction for the selected assets and the composition of the global minimum variance portfolio.

The rest of this paper is set up as follows. Section 2 discusses the methodological framework, including the construction of the time-varying rotation matrix and the specification of the dynamics of the rotation angles. Section 3 discusses the asymptotic properties of the proposed DSR model and establishes our new identification result for the rotation angles. Section 4 presents a simulation study. Section 5 applies the DSR model to US financial equities. Section 6 concludes. All proofs are collected in the Online Appendix.

Throughout, we adopt the following notational convention. Vectors and matrices are denoted in boldface, whereas scalars are non-bold. For a vector $\boldsymbol{x} = (x_1, \ldots, x_n)^\top$, its p-norm is defined by $\|\boldsymbol{x}\|_p = (\sum_{j=1}^n |x_j|^p)^{1/p}$. The induced p-norm of a matrix $\boldsymbol{A}$ is given by $\|\boldsymbol{A}\|_p = \sup_{\boldsymbol{x} \neq \boldsymbol{0}} \|\boldsymbol{A}\boldsymbol{x}\|_p / \|\boldsymbol{x}\|_p$. When $p = 2$, the subscript is omitted throughout. Moreover, let $\boldsymbol{A}^{\odot 2} = \boldsymbol{A} \odot \boldsymbol{A}$, where $\odot$ denotes the Hadamard product, and let $\det(\cdot)$ denote the determinant of a square matrix. A block-diagonal matrix with, for instance, two blocks $\boldsymbol{A}$ and $\boldsymbol{B}$ is denoted by $\operatorname{diag}(\boldsymbol{A}, \boldsymbol{B})$. A vector-valued sequence $\{\boldsymbol{a}_t,\ t \in \mathbb{Z}\}$ is said to converge to zero exponentially fast almost surely (e.a.s.) as $t \to \infty$, denoted by $\boldsymbol{a}_t \xrightarrow{e.a.s.} \boldsymbol{0}$, if there exists some $\gamma > 1$ such that $\gamma^t \|\boldsymbol{a}_t\| \xrightarrow{a.s.} 0$, where $\xrightarrow{a.s.}$ denotes almost sure (a.s.) convergence.

2 Methodology

This section introduces the proposed methodological framework. Section 2.1 presents the spectral modeling setup with the orthogonal rotation scheme aimed to capture the dynamics of the time-varying eigenvectors. Section 2.2 then discusses the score-driven dynamics of the rotation angles and likelihood-based estimation.

2.1 Spectral decomposition and eigenvector rotations

Let $\boldsymbol{y}_t \in \mathbb{R}^k$ be a vector-valued time series for $t \in \mathbb{Z}$. We assume that $\boldsymbol{y}_t \mid \mathscr{F}_{t-1}$ has a conditional pdf $p(\boldsymbol{y}_t \mid \boldsymbol{f}_t, \boldsymbol{\nu})$, where $\mathscr{F}_{t-1} = \sigma\big(\boldsymbol{y}_s,\ s \leqslant t-1\big)$ is the σ-field generated by the past data, $\boldsymbol{f}_t$ is a vector of time-varying parameters, and $\boldsymbol{\nu}$ is a vector of static parameters. We focus on the dynamics of the conditional covariance matrix $\boldsymbol{\Sigma}_t$ of $\boldsymbol{y}_t \mid \mathscr{F}_{t-1}$, which we assume always exists. Using the spectral or eigenvalue-eigenvector decomposition of $\boldsymbol{\Sigma}_t$, we write $\boldsymbol{\Sigma}_t = \boldsymbol{V}_t \boldsymbol{\Lambda}_t \boldsymbol{V}_t^\top$. We allow both the diagonal eigenvalue matrix $\boldsymbol{\Lambda}_t$ and the orthogonal eigenvector matrix $\boldsymbol{V}_t$ (and thus also $\boldsymbol{\Sigma}_t$ itself) to vary over time as functions of $\boldsymbol{f}_t$, i.e., $\boldsymbol{V}_t = \boldsymbol{V}(\boldsymbol{f}_t)$, $\boldsymbol{\Lambda}_t = \boldsymbol{\Lambda}(\boldsymbol{f}_t)$, and $\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}(\boldsymbol{f}_t)$ for some $\boldsymbol{V}(\cdot)$, $\boldsymbol{\Lambda}(\cdot)$, and $\boldsymbol{\Sigma}(\cdot)$, respectively.

The spectral representation separates the covariance dynamics into changes in scale and changes in direction. The eigenvalue matrix $\boldsymbol{\Lambda}_t$ determines the conditional variances along the orthogonal directions encoded by $\boldsymbol{V}_t$, whereas $\boldsymbol{V}_t$ determines the orientation of these directions. If $\boldsymbol{V}_t$ is kept fixed over time as in the λ-GARCH model of Hetland et al. (2023), all the time series variation in $\boldsymbol{\Sigma}_t$ is restricted to changes in scale along a fixed set of orthogonal directions. Such a restriction provides a useful benchmark, but may be restrictive. For instance, one can conceive situations in which not only the dominant eigenvalue (denoted by $\lambda_{1,t}$ below) varies over time, but also changes in the corresponding dominant eigenvector can contribute to the model's fit. Therefore, we generalize the spectral modeling approach by also allowing the orthogonal directions themselves to evolve over time. The challenge is to introduce such directional dynamics without violating the orthonormality constraints required by the spectral decomposition of a real, symmetric matrix. For this reason, we model the time variation in $\boldsymbol{V}_t$ through orthogonal rotations. For expositional purposes, we concentrate in this section only on the first eigenvector associated with $\lambda_{1,t}$.

Let $\boldsymbol{\Lambda}_t = \operatorname{diag}(\boldsymbol{\lambda}_t)$, where $\boldsymbol{\lambda}_t = (\lambda_{1,t}, \dots, \lambda_{k,t})^\top$, and specify the eigenvector matrix as $\boldsymbol{V}_t = \big(\boldsymbol{v}_{1,t}, \boldsymbol{v}_{2,t}, \dots, \boldsymbol{v}_{k,t}\big) = \bar{\boldsymbol{V}}\boldsymbol{S}_t$, where $\bar{\boldsymbol{V}}$ is a prespecified orthogonal reference matrix ($\bar{\boldsymbol{V}}^\top\bar{\boldsymbol{V}} = \bar{\boldsymbol{V}}\bar{\boldsymbol{V}}^\top = \boldsymbol{I}_k$), such as the matrix of eigenvectors of the unconditional covariance matrix, and

$$\boldsymbol{S}_t = \boldsymbol{S}(\boldsymbol{\varphi}_t) = \boldsymbol{R}_{2,1}(\varphi_{2,t})\boldsymbol{R}_{3,1}(\varphi_{3,t})\cdots\boldsymbol{R}_{k,1}(\varphi_{k,t}) = \prod_{i=2}^{k}\boldsymbol{R}_{i,1}(\varphi_{i,t}), \tag{2.1}$$

where $\boldsymbol{\varphi}_t = (\varphi_{2,t}, \dots, \varphi_{k,t})^\top$ collects all the rotation angles for the first eigenvector. For $i = 2, \dots, k$, the matrices $\boldsymbol{R}_{i,1}(\varphi)$ are the Givens rotations in the $(1, i)$ plane, defined by

$$\boldsymbol{R}_{i,1}(\varphi) = \boldsymbol{I}_k + \big[\cos(\varphi) - 1\big](\boldsymbol{E}_{ii} + \boldsymbol{E}_{11}) + \sin(\varphi)(\boldsymbol{E}_{1i} - \boldsymbol{E}_{i1}), \tag{2.2}$$

where $\boldsymbol{E}_{ij}$ denotes a $k \times k$ matrix of zeros, with a single 1 in element (i, j); see also Pinheiro and Bates (1996).

The construction in (2.1) should be interpreted as rotating eigenvectors jointly rather than separately. Each rotation $\boldsymbol{R}_{i,1}(\varphi_{i,t})$ acts only in the plane spanned by the first and the ith eigendirection. Thus, changing $\varphi_{i,t}$ tilts the leading eigendirection towards, or away from, the current ith direction, while keeping the full matrix $\boldsymbol{S}_t$ orthogonal. By compounding the rotations in the $(i, 1)$-planes for $i = 2, \dots, k$, the leading eigenvector can move freely over time in any direction, while the remaining eigenvectors move with it in a way that preserves the orthonormality of the entire eigenvector matrix $\boldsymbol{V}_t$.

To explain the construction, consider a three-dimensional ($k = 3$) example, where we omit the time index t for ease of exposition. The above rotation scheme

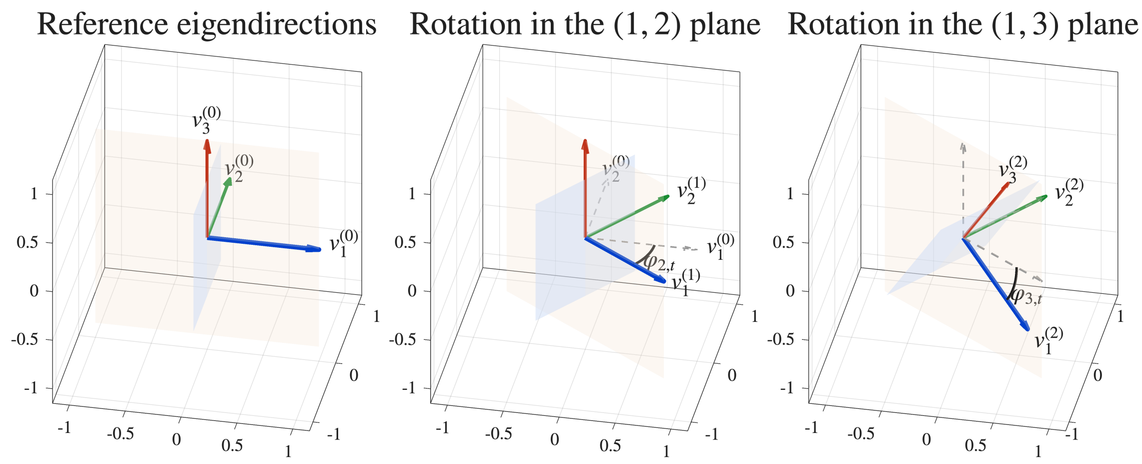

Figure 1: Illustration of the rotation (2.1) of time-varying eigenvectors in the three-dimensional case. The rotation changes the leading eigenvector direction while updating the remaining directions so that the eigenvector matrix remains orthonormal.

can then be viewed as changing the direction of the leading axis ($\boldsymbol{v}_1$) within the (current) $(i, 1)$ plane, while keeping the ith axis orthogonal to it, and leaving the remaining axis untouched. This is illustrated in Figure 1. Starting from an initial set of eigendirections $(\boldsymbol{v}_1^{(0)}, \boldsymbol{v}_2^{(0)}, \boldsymbol{v}_3^{(0)})$ in the leftmost figure, the first rotation tilts the leading eigendirection within the plane spanned by $\boldsymbol{v}_1^{(0)}$ and $\boldsymbol{v}_2^{(0)}$, producing an intermediate updated leading direction $\boldsymbol{v}_1^{(1)}$ with correspondingly updated perpendicular $\boldsymbol{v}_2^{(1)}$, while leaving $\boldsymbol{v}_3^{(1)} := \boldsymbol{v}_3^{(0)}$ untouched. The second rotation, plotted in the rightmost panel, then tilts this intermediate direction within the corresponding $(1, 3)$ plane spanned by $\boldsymbol{v}_1^{(1)}$ and $\boldsymbol{v}_3^{(1)}$, yielding the final time-varying eigendirection $\boldsymbol{v}_1^{(2)}$ with corresponding $\boldsymbol{v}_3^{(2)}$, while leaving $\boldsymbol{v}_2^{(2)} := \boldsymbol{v}_2^{(1)}$ untouched. At each step, the remaining two eigendirections are thus not updated independently, but are carried forward along the same orthogonal transformation, thereby retaining their perpendicularity to the leading eigendirection and to each other. The rotations can thus be viewed as tilting the leading eigendirection together with its orthogonal complement, rather than as updating the eigenvectors one by one. Generalizing the logic from the three-dimensional to the k-dimensional context, we have that the product of rotation matrices defining $\boldsymbol{S}_t$ only uses rotations in planes involving the first coordinate and one of the remaining coordinates $i = 2, \ldots, k$. This yields a total of $k-1$ rotation angles.

It is clear that the rotation scheme can also be generalized to eigendirections beyond the first one using rotation matrices $\boldsymbol{R}_{i,j}(\varphi_{i,j,t})$ for $i > j$, i.e., rotations in the plane orthogonal to the *preceding* eigendirections. In Section 5, however, we find that such additional flexibility brings limited gains for our empirical application, while substantially complicating the theoretical analysis. Therefore, we concentrate the theoretical development for the new model in the current paper on time variation in the first eigenvector only, as this is empirically most relevant and theoretically most transparent.

2.2 The Dynamic Spectral Rotation (DSR) model

Having introduced the rotation scheme, we now specify the dynamics of the time-varying eigenvalues and rotation angles. We stack both time-varying entities into a vector $\boldsymbol{f}_t = \left(\boldsymbol{\lambda}_t^\top, \boldsymbol{\varphi}_t^\top\right)^\top \in \mathbb{F}$, where $\mathbb{F} := \mathbb{F}_\lambda \times \mathbb{F}_\varphi$. We specify the dynamics of $\boldsymbol{f}_t$ through a score-driven update (Creal et al., 2013; Harvey, 2013), whereby the lagged value of $\boldsymbol{f}_t$ is updated in the direction of the scaled score of the conditional log-likelihood. We refer to this as the Dynamic Spectral Rotation (DSR) model. It is well established that score-driven parameter updates yield expected improvements in Kullback-Leibler divergence (Blasques et al., 2015; Creal et al., 2024; Gorgi et al., 2024) and deliver path consistency even under severe model misspecification (Beutner et al., 2026). Specifically, we consider a stochastic recurrence equation (SRE)

$$\boldsymbol{f}_{t+1} = \boldsymbol{\psi}_t(\boldsymbol{f}_t, \boldsymbol{\theta}) = (\boldsymbol{I}_{2k-1} - \boldsymbol{B})\,\boldsymbol{\omega} + \boldsymbol{B}\boldsymbol{f}_t + \boldsymbol{A}\,\boldsymbol{\mathcal{S}}(\boldsymbol{f}_t, \boldsymbol{\nu})\,\frac{\partial \ell_t(\boldsymbol{f}_t, \boldsymbol{\nu})}{\partial \boldsymbol{f}_t}, \tag{2.3}$$

where $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ collects all static parameters in $(\boldsymbol{\nu}, \boldsymbol{\omega}, \boldsymbol{B}, \boldsymbol{A})$, $\boldsymbol{\mathcal{S}}(\boldsymbol{f}_t, \boldsymbol{\nu})$ is a scaling matrix, and $\ell_t(\boldsymbol{f}, \boldsymbol{\nu}) = \log p(\boldsymbol{y}_t \mid \boldsymbol{f}, \boldsymbol{\nu})$. Throughout, we set $\boldsymbol{\mathcal{S}}(\boldsymbol{f}_t, \boldsymbol{\nu}) = \mathrm{diag}(2\boldsymbol{\Lambda}_t^2, \boldsymbol{I}_{k-1})$. This simplifies the scaling to operate separately on the eigenvalues and rotation angles, respectively, and allows us to fully develop the theoretical properties of the model.

In particular, the unit scaling choice for the rotation angles is much simpler than the inverse information matrix scaling in this case (see Creal et al., 2013), but still allows us to ensure filter invertibility. Similarly, the scaling by $2\boldsymbol{\Lambda}_t^2$ resembles inverse information matrix scaling for the eigenvalues, mimicking the common scaling choice for univariate score-driven models for variances (e.g., Creal et al., 2011, 2013). This makes sense once we realize that the eigenvalues $\lambda_{i,t}$ are exactly the variances along the ith principal directions $\boldsymbol{v}_{i,t}$. We also assume that $\boldsymbol{A} = \operatorname{diag}(\boldsymbol{A}_\lambda, \boldsymbol{A}_\varphi)$ and $\boldsymbol{B} = \operatorname{diag}(\boldsymbol{B}_\lambda, \boldsymbol{B}_\varphi)$ are block-diagonal, such that there are no spillovers between the angles and the eigenvalues, and write $\boldsymbol{\omega} = (\boldsymbol{\omega}_\lambda^\top, \boldsymbol{\omega}_\varphi^\top)^\top$.

To make the score-driven dynamics (2.3) explicit, consider for example two common choices of conditional density specifications such as the Gaussian and Student's t densities (see, e.g., Creal et al., 2011; Hetland et al., 2023). Given that $\big[\det(\boldsymbol{V}_t)\big]^2 = 1$ by the orthogonality of $\boldsymbol{V}_t$ for all $t \in \mathbb{Z}$, we obtain

$$\ell_t(\boldsymbol{f}_t, \boldsymbol{\nu}) = -\frac{k}{2}\log(2\pi) - \frac{1}{2}\log\det(\boldsymbol{\Lambda}_t) - \frac{1}{2}\boldsymbol{y}_t^\top \boldsymbol{V}_t \boldsymbol{\Lambda}_t^{-1} \boldsymbol{V}_t^\top \boldsymbol{y}_t, \tag{2.4}$$

if $\boldsymbol{y}_t \mid \mathscr{F}_{t-1} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_t)$, where $\boldsymbol{\nu}$ is an empty vector, and

$$\begin{aligned}\ell_t(\boldsymbol{f}_t, \boldsymbol{\nu}) = \log\Gamma\left(\frac{\nu+k}{2}\right) - \log\Gamma\left(\frac{\nu}{2}\right) - \frac{k}{2}\log\big((\nu-2)\pi\big) \\ - \frac{1}{2}\log\det(\boldsymbol{\Lambda}_t) - \frac{\nu+k}{2}\log\Big(1 + (\nu-2)^{-1}\boldsymbol{y}_t^\top \boldsymbol{V}_t \boldsymbol{\Lambda}_t^{-1} \boldsymbol{V}_t^\top \boldsymbol{y}_t\Big),\end{aligned} \tag{2.5}$$

if $\boldsymbol{y}_t \mid \mathscr{F}_{t-1} \sim t_\nu\big(\boldsymbol{0}, \frac{\nu-2}{\nu}\boldsymbol{\Sigma}_t\big)$, with $\boldsymbol{\nu} = \nu$.

For convenience, we introduce the following additional notation. For $i = 2, \ldots, k$, define the derivative operator $\mathscr{D}_i := \partial / \partial\varphi_i$, acting component-wise on matrix-valued functions of $\boldsymbol{\varphi} = (\varphi_2, \ldots, \varphi_k)^\top$. This immediately implies

$$\mathscr{D}_i \boldsymbol{S}(\boldsymbol{\varphi}) = \left(\prod_{\ell=2}^{i-1} \boldsymbol{R}_{\ell,1}(\varphi_\ell)\right) \boldsymbol{R}_{i,1}\left(\varphi_i + \frac{\pi}{2}\right)(\boldsymbol{E}_{11} + \boldsymbol{E}_{ii})\left(\prod_{\ell=i+1}^{k} \boldsymbol{R}_{\ell,1}(\varphi_\ell)\right). \tag{2.6}$$

That is, the operator $\mathscr{D}_i$ acts by rotating the matrix $\boldsymbol{S}(\boldsymbol{\varphi})$ by an additional $\pi/2$ angle in the $(i,1)$ plane. We also define $\boldsymbol{y}_t^{\odot 2} = \boldsymbol{y}_t \odot \boldsymbol{y}_t$, where $\odot$ is the element-wise Hadamard product. For the Gaussian and Student's t example densities in Eqs. (2.4) and (2.5), we can now formulate the following proposition.

Proposition 1. *For the DSR model, we have* $\partial \ell_t(\boldsymbol{f}_t, \boldsymbol{\nu})/\partial \boldsymbol{\lambda}_t = \frac{1}{2}\boldsymbol{\Lambda}_t^{-2}\big[w_t(\boldsymbol{V}_t^\top \boldsymbol{y}_t)^{\odot 2} - \boldsymbol{\lambda}_t\big]$ *and*

$$\frac{\partial \ell_t(\boldsymbol{f}_t, \boldsymbol{\nu})}{\partial \boldsymbol{\varphi}_t} = -w_t \begin{pmatrix} \boldsymbol{y}_t^\top \bar{\boldsymbol{V}}[\mathscr{D}_2 \boldsymbol{S}(\boldsymbol{\varphi}_t)]\boldsymbol{\Lambda}_t^{-1}\boldsymbol{V}_t^\top \boldsymbol{y}_t \\ \vdots \\ \boldsymbol{y}_t^\top \bar{\boldsymbol{V}}[\mathscr{D}_k \boldsymbol{S}(\boldsymbol{\varphi}_t)]\boldsymbol{\Lambda}_t^{-1}\boldsymbol{V}_t^\top \boldsymbol{y}_t \end{pmatrix}, \tag{2.7}$$

where $w_t := w_t(\nu) = \frac{\nu+k}{\nu-2}\Big(1 + (\nu-2)^{-1}\boldsymbol{y}_t^\top \boldsymbol{V}_t \boldsymbol{\Lambda}_t^{-1}\boldsymbol{V}_t^\top \boldsymbol{y}_t\Big)^{-1}$ *if* $\boldsymbol{y}_t \mid \mathscr{F}_{t-1} \sim t_\nu\big(\mathbf{0}, \frac{\nu-2}{\nu}\boldsymbol{\Sigma}_t\big)$, *and* $w_t = 1$ *if* $\boldsymbol{y}_t \mid \mathscr{F}_{t-1} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_t)$.

The Gaussian case is naturally nested in the Student's t case, as $w_t \to 1$ as $\nu \to \infty$. When $\nu < \infty$, the weight w_t decreases as the quadratic form $\boldsymbol{y}_t^\top \boldsymbol{V}_t \boldsymbol{\Lambda}_t^{-1}\boldsymbol{V}_t^\top \boldsymbol{y}_t$ becomes large. Hence, large realizations of $\boldsymbol{y}_t$ have a smaller impact on the update of $\boldsymbol{f}_t$, providing a natural robustness feature of the Student's t specification relative to the Gaussian specification; see also Creal et al. (2011).

Under the dynamics in (2.3), we are now ready to construct the quasi maximum likelihood estimator (QMLE) of $\boldsymbol{\theta}$. For $t \geqslant 2$, let $\hat{\boldsymbol{f}}_t(\boldsymbol{\theta}) := \hat{\boldsymbol{f}}_t(\boldsymbol{\theta}, \hat{\boldsymbol{f}}_1)$ denote the sequence initialized with a nonrandom value $\hat{\boldsymbol{f}}_1 \in \mathbb{F}$. The QMLE of the static parameters is then defined as

$$\widehat{\boldsymbol{\theta}}_T = \widehat{\boldsymbol{\theta}}_T\big(\hat{\boldsymbol{f}}_1\big) \in \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\operatorname{argmax}}\, \widehat{\mathcal{L}}_T(\boldsymbol{\theta}) = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\operatorname{argmax}}\, T^{-1}\sum_{t=1}^{T} \ell_t\big(\hat{\boldsymbol{f}}_t(\boldsymbol{\theta}), \boldsymbol{\nu}\big). \tag{2.8}$$

The angles in the vector $\boldsymbol{\varphi}_t$ are allowed to evolve freely along the real line. Given the periodicity of the Givens rotation matrices $\boldsymbol{R}_{i,1}(\varphi_{i,t})$, it is tempting to think that the angles in $\boldsymbol{\varphi}_t$ cannot be uniquely identified, except possibly over some small range such as $(-\pi/2, \pi/2]$; see also the arguments raised in Hetland et al. (2023). Somewhat surprisingly, however, we show in the next section that the model is uniquely identified under weak conditions without imposing restrictions on the filter space $\mathbb{F}_\varphi$ of $\boldsymbol{\varphi}_t$, thus allowing the angles to vary over much wider ranges.

3 Asymptotic properties of the QMLE

In this section we derive the asymptotic properties of the QMLE in (2.8) under the Gaussian model specification (2.4). The process $\boldsymbol{y}_t$ itself, however, is not assumed to be Gaussian, and the results in this section can therefore be interpreted within a quasi-likelihood framework. Section 3.1 first derives stationarity and ergodicity properties of the data generating process and establishes filter invertibility. Section 3.2 then studies point identification of the static parameters, with particular attention to the additional observational equivalences induced by the rotation-based eigenvector dynamics. Finally, Section 3.3 uses these results to establish strong consistency and asymptotic normality of the QMLE. Throughout this section, we treat $\bar{\boldsymbol{V}}$ as fixed and comment on its estimation in Section 5.

3.1 Stationarity and filter invertibility

The first step is to show that the proposed dynamic specification can generate a well-defined stationary and ergodic (SE) data generating process (DGP). The following proposition establishes this result by first deriving the stationary solution for the eigenvalue dynamics and then combining it with the rotation dynamics for the eigenvectors. This provides the probabilistic foundation for the invertibility and asymptotic results developed further below.

Proposition 2 (Existence of SE DGP)**.** *Let $\{\boldsymbol{\varepsilon}_t,\, t \in \mathbb{Z}\}$ be an i.i.d. sequence satisfying $\mathbb{E}\,\|\boldsymbol{\varepsilon}_t\|^{2\eta} < \infty$ for some $\eta > 0$. For $t \geqslant 2$ and $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, consider the sequence $\widehat{\boldsymbol{\lambda}}_t^{\varepsilon} = \widehat{\boldsymbol{\lambda}}_t^{\varepsilon}(\boldsymbol{\theta}, \widehat{\boldsymbol{\lambda}}_1^{\varepsilon}) \in \mathbb{F}_\lambda \subset [0, +\infty)^k$, initialized at $t = 1$ with a nonrandom value*

$\widehat{\boldsymbol{\lambda}}_1^{\varepsilon} \in \mathbb{F}_\lambda$, and defined recursively by the following equation:

$$\widehat{\boldsymbol{\lambda}}_{t+1}^{\varepsilon} = (\boldsymbol{I}_k - \boldsymbol{B}_\lambda)\boldsymbol{\omega}_\lambda + \boldsymbol{B}_\lambda \widehat{\boldsymbol{\lambda}}_t^{\varepsilon} + \boldsymbol{A}_\lambda \big(\widehat{\boldsymbol{\Lambda}}_t^{\varepsilon} \boldsymbol{\varepsilon}_t^{\odot 2} - \widehat{\boldsymbol{\lambda}}_t^{\varepsilon}\big) = (\boldsymbol{I}_k - \boldsymbol{B}_\lambda)\boldsymbol{\omega}_\lambda + \boldsymbol{\Phi}_t \widehat{\boldsymbol{\lambda}}_t^{\varepsilon}, \tag{3.1}$$

where $\widehat{\boldsymbol{\Lambda}}_t^{\varepsilon} = \operatorname{diag}(\widehat{\boldsymbol{\lambda}}_t^{\varepsilon})$ and $\boldsymbol{\Phi}_t = (\boldsymbol{B}_\lambda - \boldsymbol{A}_\lambda) + \boldsymbol{A}_\lambda \operatorname{diag}(\boldsymbol{\varepsilon}_t^{\odot 2})$. Assume that the space of eigenvalue processes $\mathbb{F}_\lambda$ is complete and separable.

(i) If $\mathbb{E}\big[\log\big(\|\prod_{t=1}^{r} \boldsymbol{\Phi}_t\|\big)\big] < 0$ for some integer $r \geqslant 1$, then, for every $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, the sequence $\big\{\widehat{\boldsymbol{\lambda}}_t^{\varepsilon},\, t \in \mathbb{Z}^+\big\}$ converges exponentially fast almost surely (e.a.s.) to the unique SE sequence $\big\{\boldsymbol{\lambda}_t^{\varepsilon},\, t \in \mathbb{Z}\big\}$.

(ii) Moreover, let $\boldsymbol{y}_t^{\varepsilon} = \bar{\boldsymbol{V}} \boldsymbol{S}(\boldsymbol{\varphi}_t^{\varepsilon}) \boldsymbol{\Lambda}_t^{\varepsilon,1/2} \boldsymbol{\varepsilon}_t$, where $\boldsymbol{\Lambda}_t^{\varepsilon} = \operatorname{diag}(\boldsymbol{\lambda}_t^{\varepsilon})$, $\bar{\boldsymbol{V}}$ is orthogonal, $\boldsymbol{S}(\cdot)$ is defined in (2.1), and the sequence $\{\boldsymbol{\varphi}_t^{\varepsilon},\, t \in \mathbb{Z}\}$ admits the following representation

$$\boldsymbol{\varphi}_{t+1}^{\varepsilon} = (\boldsymbol{I}_{k-1} - \boldsymbol{B}_\varphi)\boldsymbol{\omega}_\varphi + \boldsymbol{B}_\varphi \boldsymbol{\varphi}_t^{\varepsilon} + \boldsymbol{A}_\varphi \boldsymbol{s}_t^{\varepsilon,\varphi}(\boldsymbol{f}_t^{\varepsilon}, \boldsymbol{\theta}), \tag{3.2}$$

$$\boldsymbol{s}_t^{\varepsilon,\varphi}(\boldsymbol{f}_t^{\varepsilon}, \boldsymbol{\theta}) = -\begin{pmatrix} \boldsymbol{\varepsilon}_t^\top \boldsymbol{\Lambda}_t^{\varepsilon,1/2} \boldsymbol{S}(\boldsymbol{\varphi}_t^{\varepsilon})^\top [\mathscr{D}_2 \boldsymbol{S}(\boldsymbol{\varphi}_t^{\varepsilon})] \boldsymbol{\Lambda}_t^{\varepsilon,-1/2} \boldsymbol{\varepsilon}_t \\ \vdots \\ \boldsymbol{\varepsilon}_t^\top \boldsymbol{\Lambda}_t^{\varepsilon,1/2} \boldsymbol{S}(\boldsymbol{\varphi}_t^{\varepsilon})^\top [\mathscr{D}_k \boldsymbol{S}(\boldsymbol{\varphi}_t^{\varepsilon})] \boldsymbol{\Lambda}_t^{\varepsilon,-1/2} \boldsymbol{\varepsilon}_t \end{pmatrix}, \tag{3.3}$$

with $\boldsymbol{f}_t^{\varepsilon} = \binom{\boldsymbol{\lambda}_t^{\varepsilon}}{\boldsymbol{\varphi}_t^{\varepsilon}}$ for $t \in \mathbb{Z}$. If $\mathbb{E}\big(\|\prod_{t=1}^{r} \boldsymbol{\Phi}_t\|^{\eta}\big) < 1$ for some integer $r \geqslant 1$, $\mathbb{F}_\lambda \subset [\lambda_L, +\infty)^k$ for some $\lambda_L > 0$, and $\mathbb{E}\log\Big(\|\boldsymbol{B}_\varphi\| + \sqrt{2}(k-1)\lambda_L^{-1/2}\|\boldsymbol{A}_\varphi\|\|\boldsymbol{\lambda}_t^{\varepsilon}\|^{1/2}\|\boldsymbol{\varepsilon}_t\|^2\Big) < 0$, then the sequence $\big\{\boldsymbol{y}_t^{\varepsilon},\, t \in \mathbb{Z}\big\}$ is SE with $\mathbb{E}\,\|\boldsymbol{y}_t^{\varepsilon}\|^{2\eta} < \infty$.

Several remarks are in order. First, Proposition 2 does not impose a distributional assumption on the innovation sequence $\boldsymbol{\varepsilon}_t$: it only requires the moment condition stated in the proposition. Second, we do not establish the contraction of the *joint* initialized sequence $(\widehat{\boldsymbol{\lambda}}_t^{\varepsilon}, \widehat{\boldsymbol{\varphi}}_t^{\varepsilon})$, since such an argument would require restrictive uniform boundedness conditions on the eigenvalue process $\widehat{\boldsymbol{\lambda}}_t^{\varepsilon}$ in order to obtain contractivity of $\widehat{\boldsymbol{\varphi}}_t^{\varepsilon}$. Instead, Proposition 2 first shows that $\widehat{\boldsymbol{\lambda}}_t^{\varepsilon}$ admits a unique SE limit. Given this limit, the rotation angle recursion in (3.2) can be viewed as an SRE driven by the SE input process $\{(\boldsymbol{\lambda}_t^{\varepsilon}, \boldsymbol{\varepsilon}_t),\, t \in \mathbb{Z}\}$. The contraction condition $\mathbb{E}\log\Big(\|\boldsymbol{B}_\varphi\| + \sqrt{2}(k-1)\lambda_L^{-1/2}\|\boldsymbol{A}_\varphi\|\|\boldsymbol{\lambda}_t^{\varepsilon}\|^{1/2}\|\boldsymbol{\varepsilon}_t\|^2\Big) < 0$ in Part (ii) then guarantees the existence of a unique SE solution $\{\boldsymbol{\varphi}_t^{\varepsilon},\, t \in \mathbb{Z}\}$, and the corresponding initialized recursion converges to it e.a.s. Together, these results imply that $\boldsymbol{y}_t^{\varepsilon}$ is SE. Third, the moment conditions involving $\boldsymbol{\Phi}_t$, such as $\mathbb{E}\big[\log\big(\|\prod_{t=1}^{r}\boldsymbol{\Phi}_t\|\big)\big] < 0$ in Part (i), are reminiscent of the top Lyapunov exponent. Fourth, Part (ii) requires the eigenvalue processes to be bounded away from zero. Imposing $(\boldsymbol{I}_k - \boldsymbol{B}_\lambda)\boldsymbol{\omega}_\lambda \geqslant \lambda_L\, \mathbf{1}_k$ componentwise, together with componentwise nonnegativity of the matrices $\boldsymbol{A}_\lambda$ and $\boldsymbol{B}_\lambda - \boldsymbol{A}_\lambda$, suffices to ensure that all eigenvalues are bounded away from zero for all $t \in \mathbb{Z}$.

Next, we impose conditions ensuring the invertibility of the score-driven filter.

Assumption 1 (Invertibility). *(i) Assume that $\boldsymbol{\Theta}$ is compact, and that the filter space $\mathbb{F} := \mathbb{F}_\lambda \times \mathbb{F}_\varphi \subset [\lambda_L, +\infty)^k \times \mathbb{R}^{k-1}$ is convex, complete, and separable for some $\lambda_L > 0$.*

(ii) Let $\{\boldsymbol{y}_t,\, t \in \mathbb{Z}\}$ be an SE sequence with $\mathbb{E}\big(\log^+ \|\boldsymbol{y}_t\|\big) < \infty$. There exists an $\alpha > 0$ such that the following conditions hold: $\mathbb{E}\log\big(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\big\{\|\boldsymbol{B}_\lambda - \boldsymbol{A}_\lambda\| + \alpha\,\|\boldsymbol{A}_\varphi\|\,\lambda_L^{-2}\sqrt{k-1}\,\|\boldsymbol{y}_t\|^2\big\}\big) < 0$ and $\mathbb{E}\log\big(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\big\{\|\boldsymbol{B}_\varphi\| + 2\,\|\boldsymbol{A}_\varphi\|\,\lambda_L^{-1}(k-1)\,\|\boldsymbol{y}_t\|^2 + 2\,\alpha^{-1}\,\|\boldsymbol{A}_\lambda\|\,\sqrt{k-1}\,\|\boldsymbol{y}_t\|^2\big\}\big) < 0.$

Assumption 1(i) is common and mild. Assumption 1(ii) only requires the process $\{\boldsymbol{y}_t,\, t \in \mathbb{Z}\}$ to be SE and does not impose any distributional assumption. The remaining two logarithmic moment conditions in Assumption 1 are sufficient contraction conditions for the joint filter. The first condition controls the contraction of the eigenvalue block, while the second controls the contraction of the rotation angle block. The additional terms involving $\|\boldsymbol{y}_t\|^2$ arise from the dependence between the eigenvalue and eigenvector scores: changes in $\boldsymbol{\lambda}_t$ affect the update of the rotation angles, and changes in $\boldsymbol{\varphi}_t$ affect the update of the eigenvalues. The constant $\alpha > 0$ is the weight used in the block norm for $\boldsymbol{f}_t = (\boldsymbol{\lambda}_t^\top, \boldsymbol{\varphi}_t^\top)^\top$. It balances the two blocks: increasing α relaxes the term multiplied by α^{-1} in the second condition, but tightens the term multiplied by α in the first condition, and conversely. Thus, the assumption requires that the two blocks can be weighted so that the joint score-driven update is contractive on average.

The following proposition establishes the invertibility of the score-driven filter.

Proposition 3 (Invertibility). *Consider the SRE given by (2.3) with the Gaussian quasi scores from Proposition 1. Under Assumption 1, there exists a unique SE sequence* $\{\boldsymbol{f}_t(\cdot),\, t \in \mathbb{Z}\}$ *and a constant* $\gamma > 1$ *such that* $\gamma^t \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\hat{\boldsymbol{f}}_t(\boldsymbol{\theta}) - \boldsymbol{f}_t(\boldsymbol{\theta})\| \xrightarrow{a.s.} 0$ *as* $t \to \infty$. *Moreover,* $\hat{\boldsymbol{f}}_t(\cdot)$ *and* $\boldsymbol{f}_t(\cdot)$ *are uniformly continuous on* $\boldsymbol{\Theta}$ *for all* $t \in \mathbb{Z}^+$ *and* $t \in \mathbb{Z}$, *respectively.*

Proposition 3 shows that the initialized sequence generated by (2.3) from the Gaussian quasi scores in Proposition 1 asymptotically "forgets" its initial value and converges exponentially fast to a unique SE limit, uniformly over the parameter space. This result allows the likelihood based on the initialized filter to be asymptotically replaced by its stationary counterpart, which is a key step for establishing the consistency of the QMLE.

3.2 Point identification

We now turn to identification of the static parameters, which is a fundamental theoretical issue for the proposed model. Identification is nontrivial because the rotation angles are not uniquely determined by the conditional covariance matrix alone. The key insight here is that the structure of the rotation matrix in (2.1), together with the dynamic evolution of $\boldsymbol{f}_t$, restores point identification under mild conditions.

Write $\boldsymbol{\omega}_\lambda = (\omega_{\lambda,1}, \ldots, \omega_{\lambda,k})^\top$ and $\boldsymbol{\omega}_\varphi = (\omega_{\varphi,2}, \ldots, \omega_{\varphi,k})^\top$. Throughout, the true parameters are denoted by an extra subscript 0. We now formulate the following assumption for parameter identification.

Assumption 2 (Identification). *(i) For every* $t \in \mathbb{Z}$ *and for almost every realization of* $\mathscr{F}_{t-1}$, *the conditional distribution of* $\boldsymbol{y}_t$ *given* $\mathscr{F}_{t-1}$ *admits a pdf that is* *strictly positive on some nonempty open subset of* $\mathbb{R}^k$*. Moreover,* $\det(\boldsymbol{A}_{\lambda 0}) \neq 0$ *and* $\det(\boldsymbol{A}_{\varphi 0}) \neq 0$*.*

(ii) The parameter space $\boldsymbol{\Theta}$ *satisfies* $\omega_{\lambda,i} > 0$ *for* $i = 1, \ldots, k$*, and* $\omega_{\lambda,1} > \omega_{\lambda,k}$*, and* $\omega_{\varphi,i} \in (\omega_L, \omega_L + \pi]$ *for* $i = 2, \ldots, k$ *and some* $\omega_L \in \mathbb{R}$*. The matrices* $\boldsymbol{B}_\lambda - \boldsymbol{A}_\lambda$ *and* $\boldsymbol{B}_\varphi$ *are diagonal, and either the first or last (or both) diagonal elements of* $\boldsymbol{B}_\lambda - \boldsymbol{A}_\lambda$ *are nonzero. Moreover,* $\|\boldsymbol{B}_\lambda\| < 1$ *and* $\|\boldsymbol{B}_\varphi\| < 1$*.*

The full rank condition on $\boldsymbol{A}_{\lambda 0}$ in Assumption 2(i) is analogous to rank conditions commonly imposed in multivariate GARCH models, such as Assumption 3.5 of Hetland et al. (2023). In the present model, however, the rotation angles that determine the eigenvectors also vary over time, so that the identification problem cannot be reduced to the eigenvalue recursion alone. We therefore separately impose a similar rank requirement on $\boldsymbol{A}_{\varphi 0}$ for the dynamics of the rotation angles. The distributional assumption on the conditional density of $\boldsymbol{y}_t$ is mild and typically satisfied for common applications. Together, these conditions ensure that the random quantities entering the updates of $\boldsymbol{\lambda}_t$ and $\boldsymbol{\varphi}_t$ contain sufficient local variation to pin down the static parameters once the rotation-induced ambiguities have been resolved.

The role of Assumption 2(ii) is then to resolve these ambiguities by removing the remaining label, sign, and branch indeterminacies induced by the rotation parametrization. The restrictions $\omega_{\lambda,1} > \omega_{\lambda,k}$ and the requirement that at least one of the first and last diagonal elements of $\boldsymbol{B}_\lambda - \boldsymbol{A}_\lambda$ be nonzero are specific to our setting, in which only the dominant eigenvector is rotated, and require an ordering only between the first and last element of $\boldsymbol{\omega}_\lambda$. This should not be interpreted as imposing a pathwise ordering of the eigenvalue processes. The eigenvalues $\lambda_{i,t}$ themselves are still allowed to cross over time. Rather, the restriction fixes the labeling of the first and last eigenvalue dynamics through their static location parameters, which correspond to their unconditional means under the correctly specified score-driven recursion.

The restrictions on $\boldsymbol{\omega}_\varphi$ similarly fix the remaining sign and periodicity indeterminacies in the rotation angle dynamics. This is different from imposing restrictions directly on the realized rotation angles. For instance, Hetland et al. (2023, Assumption 3.6) impose a compact interval restriction on the static rotation angles in their identification argument. In contrast, our rotation angles are themselves time-varying state variables, and we do not require the realized path of $\boldsymbol{\varphi}_t$ to remain in a fixed domain, i.e., the elements of $\boldsymbol{\varphi}_t$ themselves are allowed to move (far) outside, for instance, the $(0, \pi]$ range and to cross at will. The identification of the true path of dynamic angles can then be established using the domain of the static intercepts in $\boldsymbol{\omega}_\varphi$ only. Finally, the conditions $\|\boldsymbol{B}_\lambda\| < 1$ and $\|\boldsymbol{B}_\varphi\| < 1$ simply require the autoregressive coefficient matrix $\boldsymbol{B} = \operatorname{diag}(\boldsymbol{B}_\lambda, \boldsymbol{B}_\varphi)$ in the updates of $\boldsymbol{f}_t$ to lie inside the unit circle, which is a standard and mild restriction in score-driven models.

Our identification statement in Proposition 4 starts with the identification of the eigenvalue dynamics. As mentioned, identification of the eigenvector dynamics is more involved, because the rotation matrix admits discrete equivalent representations through sign changes and shifts by multiples of π. Proposition 4(ii) therefore first characterizes these equivalences and then concludes in part (iii) that these can be resolved to identify the static parameters that characterize the rotation angle dynamics. In the discussion below, whenever helpful, we make the dependence on the static parameters $\boldsymbol{\theta}$ explicit; in such cases, for instance, we write $\boldsymbol{\lambda}_t(\boldsymbol{\theta})$ and $\boldsymbol{\varphi}_t(\boldsymbol{\theta})$ instead of $\boldsymbol{\lambda}_t$ and $\boldsymbol{\varphi}_t$, respectively.

Proposition 4. *Consider the Gaussian version of the DSR model given in* (2.3) *and Proposition 1. For* $\boldsymbol{\theta}, \boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$, *suppose that* $\boldsymbol{\Sigma}_t(\boldsymbol{\theta}) = \boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0)$ *a.s. for all* $t \in \mathbb{Z}$, *and let Assumption 2 hold. Then the results below follow:*

(i) $(\boldsymbol{\omega}_\lambda, \boldsymbol{B}_\lambda, \boldsymbol{A}_\lambda) = (\boldsymbol{\omega}_{\lambda 0}, \boldsymbol{B}_{\lambda 0}, \boldsymbol{A}_{\lambda 0})$;

(ii) There exist integer vectors $\boldsymbol{q}_t, \boldsymbol{q}_{t+1} \in \mathbb{Z}^{k-1}$ *such that, a.s.,* $\boldsymbol{\varphi}_t(\boldsymbol{\theta}) = \boldsymbol{\zeta}(\boldsymbol{q}_t)\, \boldsymbol{\varphi}_t(\boldsymbol{\theta}_0) + \pi\, \boldsymbol{q}_t$ *and* $\boldsymbol{\varphi}_{t+1}(\boldsymbol{\theta}) = \boldsymbol{\zeta}(\boldsymbol{q}_{t+1})\, \boldsymbol{\varphi}_{t+1}(\boldsymbol{\theta}_0) + \pi\, \boldsymbol{q}_{t+1}$ *for every* $t \in \mathbb{Z}$, *where* $\boldsymbol{\zeta}(\boldsymbol{q}) := \operatorname{diag}\left(\zeta_2(\boldsymbol{q}), \ldots, \zeta_k(\boldsymbol{q})\right)$ *is a diagonal sign matrix with* $\zeta_2(\boldsymbol{q}) := 1$ *and*

$\zeta_i(\boldsymbol{q}) := (-1)^{\sum_{j=2}^{i-1} q_j}$ *for* $i = 3, \ldots, k$, *for any integer vector* $\boldsymbol{q} = (q_2, \ldots, q_k)^\top \in \mathbb{Z}^{k-1}$*;*

(iii) $(\boldsymbol{\omega}_\varphi, \boldsymbol{B}_\varphi, \boldsymbol{A}_\varphi) = (\boldsymbol{\omega}_{\varphi 0}, \boldsymbol{B}_{\varphi 0}, \boldsymbol{A}_{\varphi 0})$.

The proof of Proposition 4 proceeds in three steps. First, point identification of the static parameters governing the eigenvalue dynamics is established by exploiting the ordering $\omega_{\lambda,1} > \omega_{\lambda,k}$ and the special structure and ordering of the rotation matrices discussed in Remark 1 below. This also fixes the rotation angles up to sign flips and integer multiples of π. The remaining indeterminacies are resolved through the score-driven dynamic specification in (2.3) by exploiting the constraints of the parameter space, in particular the intervals for the intercepts $\omega_{\varphi,i} \in (0, \pi]$ for $i = 2, \ldots, k$. This dynamic identification strategy contrasts with Blasques et al. (2022), where identification is first obtained from the conditional density for a *fixed* value of the time-varying parameter, and the parameters governing its dynamics are identified only thereafter.

Remark 1 (Point identification)**.** The identification argument uses two separate features of the model. The first is the structured form of the rotation matrix $\boldsymbol{S}(\cdot)$ in (2.1), which uses only rotations in planes involving the first coordinate and therefore imposes a specific columnwise structure. This structure helps pin down the labeling of the eigenvalue processes by restricting the permutations of eigendirections that can be observationally equivalent. Once this labeling is fixed, the remaining indeterminacies are the sign and periodicity equivalences in the rotation angles. The second feature in the identification strategy is the score-driven recursion in (2.3), which links the rotation angle process across consecutive time points and rules out angle paths that would be equivalent at each individual (static) time point, but that would be incompatible with the dynamic update.

Regarding the first feature, $\boldsymbol{S}(\boldsymbol{\varphi})$ exhibits a distinctive columnwise zero pattern; see Lemma B.4 in the Online Appendix. Specifically, the first and last columns may be fully nonzero, whereas, for $j = 2, \ldots, k-1$, the entries in rows $j+1, j+2, \ldots, k$

of the jth column are always zero, i.e.,

$$\boldsymbol{S}(\boldsymbol{\varphi}) = \begin{pmatrix} \boldsymbol{S}(\boldsymbol{\varphi})_{1,1} & \boldsymbol{S}(\boldsymbol{\varphi})_{1,2} & \cdots & \boldsymbol{S}(\boldsymbol{\varphi})_{1,k-1} & \boldsymbol{S}(\boldsymbol{\varphi})_{1,k} \\ \boldsymbol{S}(\boldsymbol{\varphi})_{2,1} & \boldsymbol{S}(\boldsymbol{\varphi})_{2,2} & \cdots & \boldsymbol{S}(\boldsymbol{\varphi})_{2,k-1} & \boldsymbol{S}(\boldsymbol{\varphi})_{2,k} \\ \boldsymbol{S}(\boldsymbol{\varphi})_{3,1} & 0 & \cdots & \boldsymbol{S}(\boldsymbol{\varphi})_{3,k-1} & \boldsymbol{S}(\boldsymbol{\varphi})_{3,k} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{S}(\boldsymbol{\varphi})_{k,1} & 0 & \cdots & 0 & \boldsymbol{S}(\boldsymbol{\varphi})_{k,k} \end{pmatrix}.$$

Here, $\boldsymbol{S}(\boldsymbol{\varphi})_{j,\ell}$ denotes the (j,ℓ)th entry of $\boldsymbol{S}(\boldsymbol{\varphi})$ for $1 \leqslant j,\ell \leqslant k$. This pattern restricts the possible permutations and sign changes that can leave the conditional covariance matrix unchanged, making it possible to characterize the remaining angle equivalences explicitly.

To illustrate this, consider the case $k = 3$. Define for $j = 2,3$, $c_j = \cos(\varphi_j)$ and $s_j = \sin(\varphi_j)$. Fix $t \in \mathbb{Z}$, and suppose that $\boldsymbol{\Sigma}_t(\boldsymbol{\theta}) = \boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0)$, where Assumption 2(i) implies that $\lambda_{i,t}(\boldsymbol{\theta}_0) \neq \lambda_{j,t}(\boldsymbol{\theta}_0)$ a.s. for $i \neq j$. Then one can show that $\boldsymbol{S}\big(\boldsymbol{\varphi}_t(\boldsymbol{\theta})\big) = \boldsymbol{S}\big(\boldsymbol{\varphi}_t(\boldsymbol{\theta}_0)\big)\boldsymbol{D}_t\boldsymbol{\Pi}_t^\top$, where $\boldsymbol{D}_t = \operatorname{diag}(\pm1,\pm1,\pm1)$ is a signed diagonal matrix and $\boldsymbol{\Pi}_t^\top$ is a permutation matrix acting on the columns of $\boldsymbol{S}\big(\boldsymbol{\varphi}_t(\boldsymbol{\theta}_0)\big)\boldsymbol{D}_t$. Moreover,

$$\boldsymbol{S}(\boldsymbol{\varphi}) = \boldsymbol{R}_{2,1}(\varphi_2)\boldsymbol{R}_{3,1}(\varphi_3) = \begin{pmatrix} c_2c_3 & s_2 & c_2s_3 \\ -s_2c_3 & c_2 & -s_2s_3 \\ -s_3 & 0 & c_3 \end{pmatrix}.$$

We see that the $(3,2)$th element is zero. Note that $\boldsymbol{S}\big(\boldsymbol{\varphi}_t(\boldsymbol{\theta})\big)$ and $\boldsymbol{S}\big(\boldsymbol{\varphi}_t(\boldsymbol{\theta}_0)\big)\boldsymbol{D}_t$ must share the same special column-zero structure, as $s_3 \neq 0$ and $c_3 \neq 0$ a.s. due to Assumption 2(i). This restricts the admissible column permutations; for example, the second and third columns cannot be permuted. In fact, the only admissible possibilities are the identity permutation and the permutation that swaps the first and third columns. Hence, $\boldsymbol{\Pi}_t^\top = \boldsymbol{\Pi}_{13}^{\delta_t}$, where $\boldsymbol{\Pi}_{13} = \left(\begin{smallmatrix} 0&0&1\\0&1&0\\1&0&0 \end{smallmatrix}\right)$ and $\delta_t \in \{0,1\}$. As a consequence, $\boldsymbol{\lambda}_t(\boldsymbol{\theta}) = \boldsymbol{\Pi}_{13}^{\delta_t}\boldsymbol{\lambda}_t(\boldsymbol{\theta}_0)$.

Applying the same argument at time $t+1$, we get $\boldsymbol{\lambda}_{t+1}(\boldsymbol{\theta}) =$

$\boldsymbol{\Pi}_{13}^{\delta_{t+1}}\boldsymbol{\lambda}_{t+1}(\boldsymbol{\theta}_0)$ for some $\delta_{t+1} \in \{0,1\}$. Hence, we obtain $\big(\boldsymbol{\lambda}_t(\boldsymbol{\theta}), \boldsymbol{\lambda}_{t+1}(\boldsymbol{\theta})\big) = \big(\boldsymbol{\Pi}_{13}^{\delta_t}\boldsymbol{\lambda}_t(\boldsymbol{\theta}_0), \boldsymbol{\Pi}_{13}^{\delta_{t+1}}\boldsymbol{\lambda}_{t+1}(\boldsymbol{\theta}_0)\big)$ a.s. for some $(\delta_t, \delta_{t+1}) \in \{0,1\} \times \{0,1\}$. For this specific t, set $(\delta, \delta') = (\delta_t, \delta_{t+1})$. By an open-support argument, which follows from Assumption 2(i) (see Lemma B.3(i) in the Online Appendix for details), the identity $\big(\boldsymbol{\lambda}_t(\boldsymbol{\theta}), \boldsymbol{\lambda}_{t+1}(\boldsymbol{\theta})\big) = \big(\boldsymbol{\Pi}_{13}^{\delta}\boldsymbol{\lambda}_t(\boldsymbol{\theta}_0), \boldsymbol{\Pi}_{13}^{\delta'}\boldsymbol{\lambda}_{t+1}(\boldsymbol{\theta}_0)\big)$ then yields $\boldsymbol{B}_\lambda - \boldsymbol{A}_\lambda = \boldsymbol{\Pi}_{1k}^{\delta'}(\boldsymbol{B}_{\lambda 0} - \boldsymbol{A}_{\lambda 0})\boldsymbol{\Pi}_{1k}^{\delta}$ and $\boldsymbol{A}_\lambda = \boldsymbol{\Pi}_{1k}^{\delta'}\boldsymbol{A}_{\lambda 0}\boldsymbol{\Pi}_{1k}^{\delta}$. Using the diagonality of $\boldsymbol{B}_\lambda - \boldsymbol{A}_\lambda$ in Assumption 2(ii), this further forces $\delta = \delta' \in \{0,1\}$, and leads to $\boldsymbol{\omega}_\lambda = \boldsymbol{\Pi}_{13}^{\delta}\boldsymbol{\omega}_{\lambda 0}$. If we additionally impose $\omega_{\lambda,1} > \omega_{\lambda,3}$ and $\omega_{\lambda 0,1} > \omega_{\lambda 0,3}$, then the case $\delta = 1$ is ruled out, so no nontrivial permutation remains. This establishes identification of the static parameters governing the dynamics of $\boldsymbol{\lambda}_t$.

The identification of the static parameters in the dynamics of $\boldsymbol{\varphi}_t$ is more involved, but the result follows along the same line of reasoning. Specifically, the conditions in Assumption 2(ii) are used to show that the integer vectors $\boldsymbol{q}_t$ and $\boldsymbol{q}_{t+1}$ in Proposition 4(ii) satisfy $\boldsymbol{q}_t = \boldsymbol{q}_{t+1} = \mathbf{0}$; see the proof of Proposition 4 for the full argument. □

3.3 Strong consistency and asymptotic normality

The following theorem shows that the QMLE converges almost surely to the true static parameter.

Theorem 1 (Strong consistency). *Let Assumption 1 and all conditions in Proposition 4 hold. Suppose that the filter dynamics are generated by the Gaussian quasi-likelihood recursion in (2.3) and Proposition 1 with true parameter value $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$, and that $\sup_{t\in\mathbb{Z}} \mathbb{E}\,\|\boldsymbol{y}_t\|^2 < \infty$. Then $\widehat{\boldsymbol{\theta}}_T \xrightarrow{a.s.} \boldsymbol{\theta}_0$ as $T \to \infty$.*

Further notation and assumptions are required to establish the asymptotic normality of the QMLE. For $\epsilon > 0$, let $\mathbb{C}(\boldsymbol{\theta}_0, \epsilon)$ denote the open cube centered at $\boldsymbol{\theta}_0$ with edge length 2ϵ.

Assumption 3 (Asymptotic normality). *(i) Let $\boldsymbol{\theta}_0 \in \mathrm{int}(\boldsymbol{\Theta})$ be an interior point* *of* $\boldsymbol{\Theta}$.

(ii) There exists some $\epsilon > 0$ *such that* $\boldsymbol{\Theta}_0^{\mathbb{C},\epsilon} = \boldsymbol{\Theta} \cap \mathbb{C}(\boldsymbol{\theta}_0, \epsilon)$ *is nonempty and, for each* $t \in \mathbb{Z}$, *the map* $\boldsymbol{\theta} \mapsto \boldsymbol{f}_t(\boldsymbol{\theta})$ *admits continuous partial derivatives up to order two on* $\boldsymbol{\Theta}_0^{\mathbb{C},\epsilon}$. *Moreover, for each* $\boldsymbol{\theta} \in \boldsymbol{\Theta}_0^{\mathbb{C},\epsilon}$ *and each* $t \in \mathbb{Z}$, *the first- and second-order partial derivatives of* $\boldsymbol{f}_t(\boldsymbol{\theta})$ *are* $\mathscr{F}_{t-1}$*-measurable; as processes indexed by* t, *these derivatives are strictly stationary and ergodic.*

(iii) $\mathbb{E}\Big(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_0^{\epsilon}} \big\|\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\ell_t(\boldsymbol{f}_t(\boldsymbol{\theta}),\boldsymbol{\nu})\big\|\Big) < \infty$, *where* $\boldsymbol{\Theta}_0^{\epsilon} \subset \boldsymbol{\Theta} \cap \mathbb{C}(\boldsymbol{\theta}_0, \epsilon)$ *is compact. Moreover,* $\mathbb{E}\Big(\Big\|\frac{\partial\ell_t(\boldsymbol{f}_t(\boldsymbol{\theta}),\boldsymbol{\nu})}{\partial\boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\Big\|^2\Big) < \infty$ *and* $\boldsymbol{\mathcal{J}}_0 := -\mathbb{E}\Big(\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\ell_t(\boldsymbol{f}_t(\boldsymbol{\theta}),\boldsymbol{\nu})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\Big)$ *is invertible.*

Assumption 3 collects additional conditions for establishing asymptotic normality. Condition (i) is standard; see, for example, Blasques et al. (2022, Theorem 4.15). Assumption 3(ii) concerns the derivative filters. It requires the first- and second-order derivative processes of $\boldsymbol{f}_t(\cdot)$ to be SE; see, for example, Artemova (2025, Assumption 10). Condition (iii) imposes moment restrictions sufficient for applying a uniform law of large numbers to the Hessian and a central limit theorem to the score. Conditions (ii)–(iii) can, in principle, be verified from more primitive contraction and moment assumptions; see Lin and Lucas (2025, Proposition 1). We state them directly because deriving such primitive conditions for the present model would require lengthy and cumbersome calculations for the derivative filters and their moments, without providing additional theoretical insights beyond standard arguments in score-driven asymptotic theory.

The following result provides a limiting approximation for the QMLE.

Theorem 2 (Asymptotic normality). *Let* $\{\boldsymbol{y}_t, t \in \mathbb{Z}\}$ *be generated according to* $\{\boldsymbol{y}_t^{\varepsilon}, t \in \mathbb{Z}\}$ *in Proposition 2, where* $\boldsymbol{\varepsilon}_t \sim (\mathbf{0}, \boldsymbol{I}_k)$ *are i.i.d. Suppose that the conditions of Theorem 1 and Assumption 3 hold. Then* $\sqrt{T}(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}\big(\mathbf{0}, \boldsymbol{\mathcal{J}}_0^{-1}\boldsymbol{\mathcal{I}}_0\boldsymbol{\mathcal{J}}_0^{-1}\big)$ *as* $T \to \infty$, *where* $\boldsymbol{\mathcal{J}}_0$ *is defined in Assumption 3(iii), and* $\boldsymbol{\mathcal{I}}_0 = \mathbb{E}\Big(\frac{\partial\ell_t(\boldsymbol{f}_t(\boldsymbol{\theta}),\boldsymbol{\nu})}{\partial\boldsymbol{\theta}}\frac{\partial\ell_t(\boldsymbol{f}_t(\boldsymbol{\theta}),\boldsymbol{\nu})}{\partial\boldsymbol{\theta}^\top}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\Big)$.

Table 1: Static parameter accuracy

For sample sizes $T = 1000, 2000, 5000$, the table shows the Monte Carlo mean and standard deviation of the ML estimates.

		$T = 1000$		$T = 2000$		$T = 5000$	
Parameter	True	Mean	Std.	Mean	Std.	Mean	Std.
α_λ	0.1	0.100	0.016	0.100	0.011	0.100	0.007
β_λ	0.8	0.800	0.031	0.799	0.022	0.800	0.013
α_φ	0.05	0.051	0.007	0.051	0.005	0.050	0.003
β_φ	0.9	0.879	0.045	0.890	0.023	0.897	0.013
$\omega_{\varphi,2}$	1.0	1.002	0.049	1.001	0.034	1.001	0.021
$\omega_{\varphi,3}$	0.5	0.504	0.043	0.502	0.030	0.501	0.018

4 Simulations

To illustrate the theoretical results on identifiability and to see whether the asymptotic distribution provides a good approximation in finite samples, we perform several simulation experiments. First, we focus on the accuracy of the consistency and asymptotic normality results in finite samples. For this, we simulate data according to $\boldsymbol{y}_t \mid \mathscr{F}_{t-1} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_t)$, where the eigenvector and eigenvalue processes of $\boldsymbol{\Sigma}_t$ follow the dynamics in (2.3) and (2.4). Moreover, we set $k = 3$, $\boldsymbol{\omega}_\lambda = (\omega_{\lambda,1}, \omega_{\lambda,2}, \omega_{\lambda,3})^\top = (2, 1, 0.5)^\top$, $\boldsymbol{\omega}_\varphi = (\omega_{\varphi,2}, \omega_{\varphi,3})^\top = (1, 0.5)^\top$, $\boldsymbol{A}_\lambda = \alpha_\lambda \boldsymbol{I}_k$, $\boldsymbol{B}_\lambda = \beta_\lambda \boldsymbol{I}_k$, $\boldsymbol{A}_\varphi = \alpha_\varphi \boldsymbol{I}_{k-1}$, and $\boldsymbol{B}_\varphi = \beta_\varphi \boldsymbol{I}_{k-1}$, with $(\alpha_\lambda, \beta_\lambda, \alpha_\varphi, \beta_\varphi) = (0.1, 0.8, 0.05, 0.9)$. For each simulated time series, we then estimate the static parameters via (2.8).

In Table 1 we report the accuracy of the ML estimates using 1000 Monte Carlo simulations and several empirically relevant sample sizes $T = 1000, 2000, 5000$. For each parameter, we report the true value and the Monte Carlo average and standard deviation of the ML estimates. The results confirm that the ML estimator is highly accurate in finite samples, already starting from sample sizes of $T = 1000$.

Figure 2 presents kernel density estimates (KDE) of the t-statistics for the same experiments. We see that for all sample sizes considered, the t-statistics are approximately normally distributed, in line with the asymptotic normality results. Only the t statistic of β_φ seems to need a somewhat larger sample size T to truly

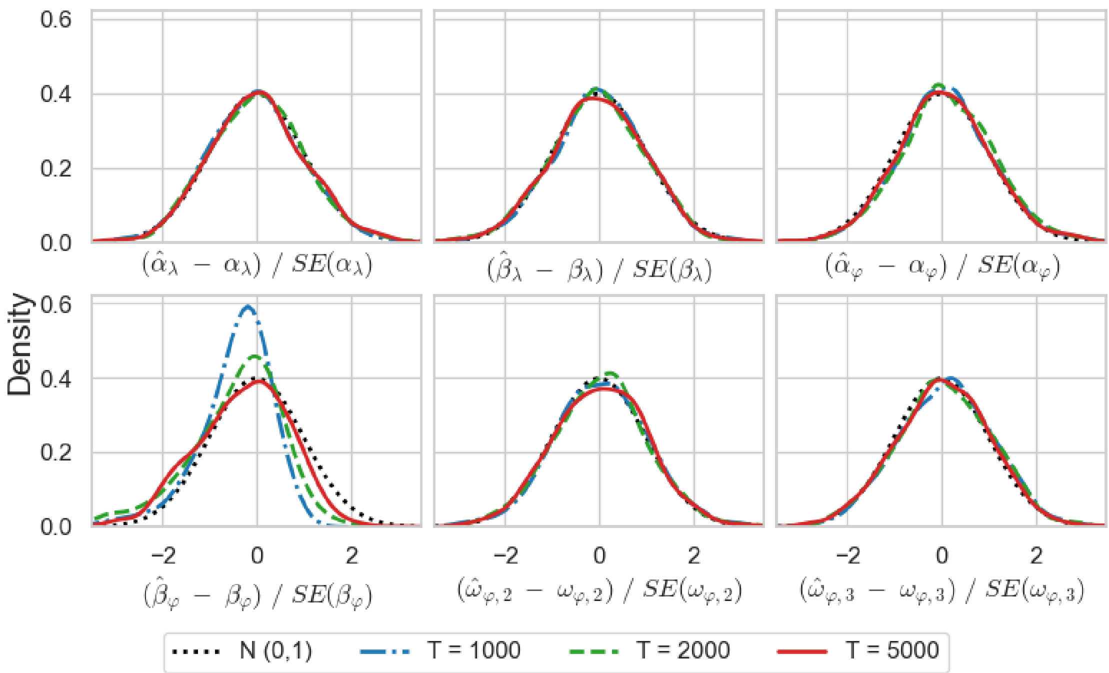

Figure 2: Empirical distribution

Kernel density estimates of the empirical distribution of each standardized estimator for different sample sizes $T \in \{1000, 2000, 5000\}$. Each estimator is standardized by centering at the true parameter value and scaling by the standard deviation of the corresponding parameter estimates across the Monte Carlo repetitions. The black dotted line represents the pdf of the standard normal distribution.

converge to the normal distribution. All other parameters, however, already seem to have converged for the smallest sample size considered in the experiment.

To illustrate our identification result, we run a second simulation experiment. For this, we simulate a path of $T = 5000$ observations and estimate the model's static parameters. Using these, we filter the time-varying eigenvalues and rotation angles. The results shown in Figure 3 clearly illustrate some of the more surprising elements of the identification result. First, tick marks are placed on the horizontal axis whenever the eigenvalues cross. We see that this happens regularly over the sample period. However, it does not impede the recovery of the eigenvalue paths by the DSR model in any way. The true and filtered paths continue to lie virtually on top of each other, also during crossings. Similarly, we see that the rotation angles can move over considerable ranges. In particular, they clearly move far outside

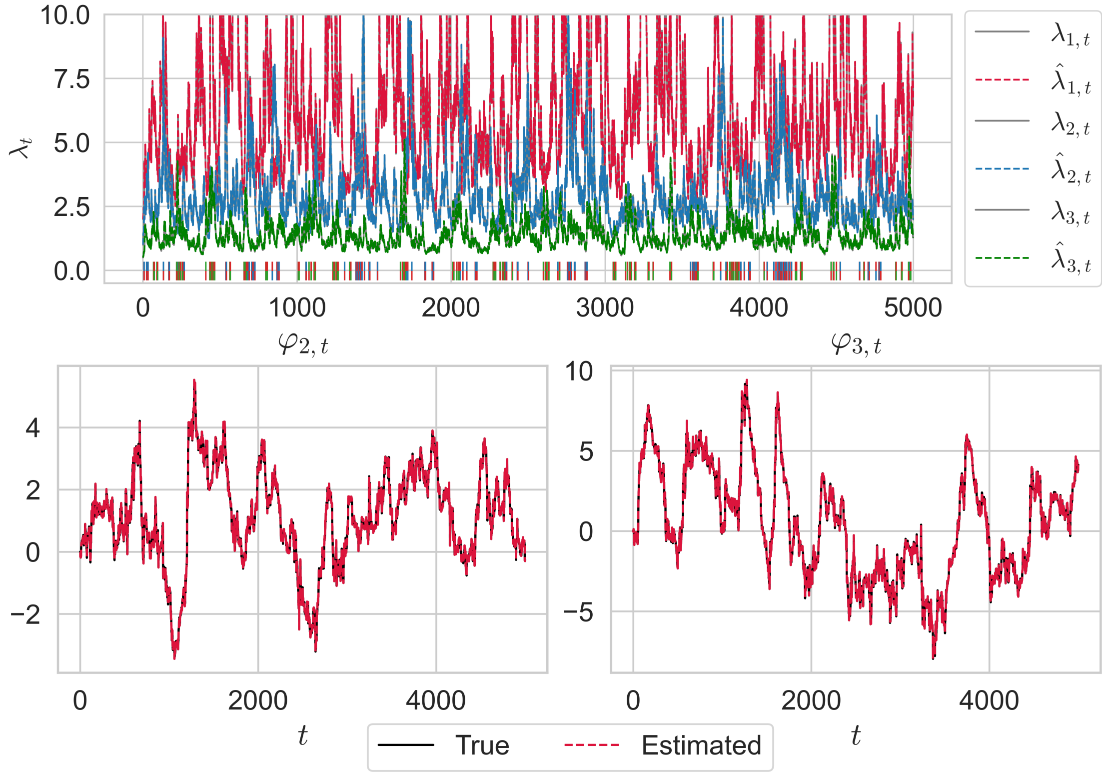

Figure 3: Identification

This figure shows true and estimated time-varying parameter paths for our second simulation experiment to assess the identifiability of the time-varying parameters.

the $(0, \pi]$ restricted ranges that hold for the intercepts $\omega_{\varphi,i}$, with for instance $\varphi_{3,t}$ ranging from around -8 to $+10$. Despite this, the estimated paths for the rotation angles lie on top of their true values for all t. So even though the high value of $\varphi_{3,t} \approx 10$ around $t = 1250$ could, for instance, in isolation be replaced by its modulo π counterpart, such a replacement would not be in line with the score-driven dynamic specification. This results in the filtered path continuing its trajectory outside the $(0, \pi]$ range, in line with the true simulated path. The novel dynamic rather than static identification strategy from Section 3.2 formally explains this finding.

5 Empirical application

This section discusses the empirical application of the DSR model to US financial equities. We first discuss the data for the application, the performance metrics, and the benchmark models in Sections 5.1, 5.2, and 5.3, respectively. Section 5.4 then presents the in-sample and out-of-sample estimation results.

5.1 Data

We apply the DSR model to daily returns on 10 financial US equities from the S&P 500 index: C, JPM, BAC, GS, WFC, KEY, COF, MTB, USB, and AXP. The underlying firms are listed in order of their SRISK systemic importance, with the most systemically important institutions listed first.[1] The dataset spans the period January 2002 to January 2023. As we are mainly interested in the predictive performance of the model, we produce one-step-ahead predictions of the fitted covariance matrices and re-estimate the model parameters approximately every month (20 days). The initial estimation window size is set to 1,250 days, spanning from January 2002 to January 2007. The out-of-sample period then contains 4,000 observations between January 2007 and January 2023. The 16-year out-of-sample window thus contains several periods of increased financial turmoil, such as the 2008 financial crisis, the European sovereign debt crisis, the 2011 US downgrade by S&P, and COVID-19, all of which put the stability and accuracy of the proposed model to the test.

5.2 Performance metrics

We compare the different models in terms of the log-scores of their density forecasts, with the time-t log-score computed as $\log p(\boldsymbol{y}_t \mid \boldsymbol{\Sigma}_t, \boldsymbol{\nu})$. The statistical significance of differences in log-scores is assessed by the Diebold-Mariano (DM) test (Diebold and Mariano, 2002). In addition, we measure model performance in terms of the

[1]For this, we take the sample averages over the entire sample period of the SRISK values of each institution as reported on `https://vlab.stern.nyu.edu/srisk/RISK.USFIN-MR.MES`.

Frobenius norm $\|\cdot\|_F$ and QLIKE (Bollerslev et al., 2018). At time $t+1$, define

$$\text{Frob}_{t+1} = \left\| \boldsymbol{RC}_{t+1} - \widehat{\boldsymbol{\Sigma}}_{t+1} \right\|_F, \qquad \text{QLIKE}_{t+1} = \log\det\left(\widehat{\boldsymbol{\Sigma}}_{t+1}\right) + \text{tr}\left(\widehat{\boldsymbol{\Sigma}}_{t+1}^{-1}\,\boldsymbol{RC}_{t+1}\right),$$

where we compare the forecast of the covariance matrix $\widehat{\boldsymbol{\Sigma}}_{t+1}$ with a realized covariance matrix $\boldsymbol{RC}_{t+1}$ based on 5-minute intraday intervals and the guidelines of (Brownlees and Gallo, 2006; Barndorff-Nielsen et al., 2009; Shephard and Sheppard, 2010). The out-of-sample Frobenius norm and QLIKE are then averaged over time and compared across models, with statistical significance again assessed using the DM test.

We also perform an economic evaluation using a Global Minimum Variance Portfolio (GMVP) framework (Chiriac and Voev, 2011). If $\boldsymbol{w}_{t+1|t}$ denotes a vector of portfolio weights constructed at time t for a portfolio from t to $t+1$, then the weight vector that minimizes the portfolio variance $\sigma^2_{pf,t+1} = \boldsymbol{w}_{t+1|t}^\top\,\widehat{\boldsymbol{\Sigma}}_{t+1}\,\boldsymbol{w}_{t+1|t}$ is given by

$$\boldsymbol{w}_{t+1|t} = \left(w_{1,t+1|t}, \ldots, w_{k,t+1|t}\right)^\top = \left(\mathbf{1}_k^\top\,\widehat{\boldsymbol{\Sigma}}_{t+1}^{-1}\,\mathbf{1}_k\right)^{-1}\left(\widehat{\boldsymbol{\Sigma}}_{t+1}^{-1}\,\mathbf{1}_k\right), \tag{5.1}$$

with $\mathbf{1}_k$ a $(k \times 1)$ vector of ones. For each point in the out-of-sample period, we compute the portfolio return $\boldsymbol{w}_{t+1|t}^\top\,\boldsymbol{y}_{t+1}$ and report its out-of-sample standard deviation. We use Hautsch et al. (2015) to additionally compare the realized portfolio variances $rc_{pf,t+1} = \boldsymbol{w}_{t+1|t}^\top\,\boldsymbol{RC}_{t+1}\,\boldsymbol{w}_{t+1|t}$ of the GMVP across models. We also report corresponding measures of portfolio turnover (DeMiguel et al., 2009) and concentration (see, e.g., Goetzmann and Kumar, 2008), namely

$$\text{TO}_t = \sum_{i=1}^{k} \left| w_{i,t+1|t} - w_{i,t|t-1}\,\frac{1+y_{i,t}}{1+\boldsymbol{w}_{t|t-1}^\top \boldsymbol{y}_t} \right|, \qquad \text{CO}_t = \left(\sum_{i=1}^{k} w_{i,t+1|t}^2\right)^{1/2},$$

where $y_{i,t}$ denotes the ith element of $\boldsymbol{y}_t$. Finally, to assess the significance of differences in ex-post variance, we apply the DM test to the series of squared GMVP returns (denoted as r_{pf}^2).

5.3 Benchmark models and baseline eigenvectors

To compare the out-of-sample performance of the new DSR model with relevant alternatives, we consider two benchmark models in an out-of-sample analysis of the DSR model. Our first obvious benchmark is the λ-GARCH model of Hetland et al. (2023). In contrast to the $k = 3$ dimensional application in Hetland et al. (2023), the $k = 10$ dimensional out-of-sample application in this section provides a more challenging setting for both models. We specify the λ-GARCH model analogously to the DSR model, using either a Gaussian or Student's t distribution, with the only difference being whether the eigenvectors are time-varying or not. The eigenvectors for the Hetland et al. (2023) model are estimated using the two-step estimation procedure of Hetland (2020), which targets the eigenvectors via the eigenvectors of the unconditional covariance matrix of the data. We use the same target as our $\bar{\boldsymbol{V}}$ in the DSR model to keep the two models as comparable as possible.

We also experimented with alternative specifications for the time-varying eigenvectors in the DSR model. For instance, we considered a specification with rotations among the first three dominant eigenvectors only, as well as a specification where not only the first, but also the second eigenvector is allowed to vary freely over time. Empirically, however, none of these alternative specifications outperforms the simple specification with only the first eigenvector time-varying. This is not entirely unexpected: for stocks, the first eigenvector by far explains the largest portion of the variance, such that it is most worthwhile to get this principal direction right. Further eigenvalues are typically an order of magnitude smaller, such that time variation in their corresponding eigenvalues has a much smaller effect on the model's fit. We therefore focus the presentation of the results on the baseline specification where only the first eigenvector varies over time. See Appendix C for more details and empirical results for the alternative specifications mentioned above.

Table 2: Full-sample DSR estimation results

This table shows the estimation results for 10 financial US equities for the Gaussian and Student's t-based DSR model. The full-sample estimation period is approximately 21 years, with 5250 observations, ranging from the 2nd of January 2002 to the 22nd of February 2023. The parameter estimates and corresponding sandwich standard errors from Theorem 2 are given in parentheses.

	α_λ	β_λ	$\alpha_\varphi(\times 10^{-3})$	β_φ	ν	log-lik	BIC
Gaussian	0.075	0.994	0.605	0.997	——	−74324	148759
	(0.008)	(0.001)	(0.058)	(0.001)			
Student's t	0.052	0.998	0.847	0.998	6.934	−71899	143917
	(0.003)	(0.000)	(0.061)	(0.000)	(0.271)		

As our second benchmark, we compare the DSR model to a standard BEKK model (Engle and Kroner, 1995). Here, we again assume that the vector of returns $\boldsymbol{y}_t$ either follows a Gaussian or Student's t distribution. The time-varying conditional covariance matrix $\boldsymbol{\Sigma}_t$ in the BEKK is defined recursively as

$$\boldsymbol{\Sigma}_{t+1} = (1 - \alpha - \beta)\boldsymbol{\Omega} + \alpha\left(\boldsymbol{y}_t\boldsymbol{y}_t^\top\right) + \beta\boldsymbol{\Sigma}_t,$$

where $\boldsymbol{\Omega}$ is estimated as the sample covariance matrix of the observed returns. With both benchmark models in place, we can now turn to the empirical results.

5.4 Results

We first present the in-sample results for the DSR model, followed by a comparison of the DSR model with the benchmark models in an out-of-sample analysis. Table 2 shows the full-sample estimation results for the new model. We see that both the eigenvalue and the eigenvector dynamics are highly persistent, with β_λ and β_φ both close to one. These results are robust to the choice between the Gaussian and the Student's t specifications. Figure 4 plots the logarithm of the eigenvalues over time. The first log-eigenvalue clearly lies between 2 and 3 points higher than the second log-eigenvalue, meaning it is between a factor $e^2 \approx 7$ and $e^3 \approx 20$ higher than the

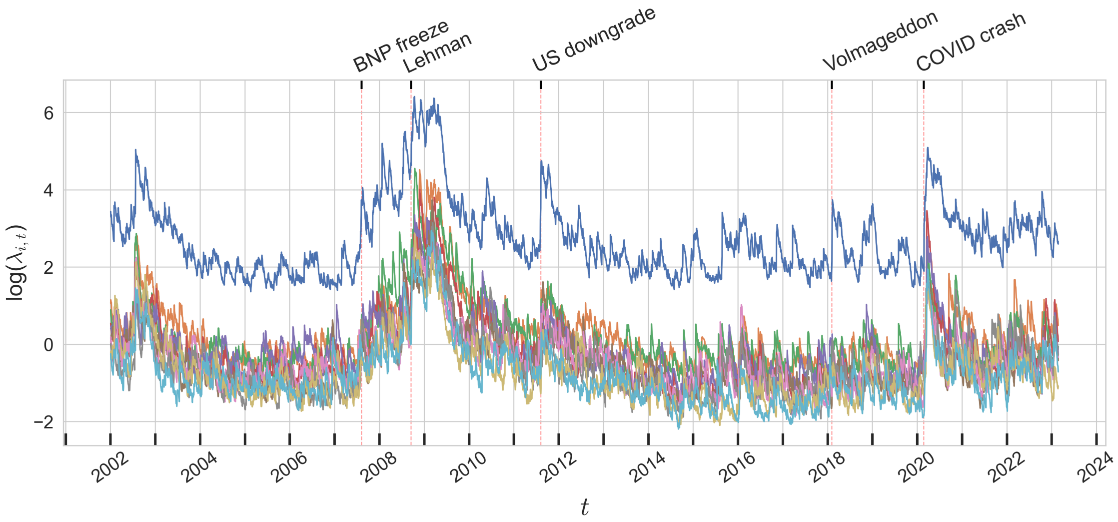

Figure 4: Dynamics of conditional eigenvalues

This figure shows the time series of all conditional log-eigenvalues. The full-sample analysis spans the period January 2002 to January 2023.

second largest eigenvalue. This is in line with the importance of a systematic risk factor such as a market or sector factor in a typical stock market analysis. The eigenvalues in Figure 4 also clearly follow the key economic events during the sample period, with volatility peaking along the principal eigendirection during major crisis times.

Figure 5 presents the time series patterns of the estimated rotation angles. All angles show modest movements over time and over similar ranges, indicating that all of these rotations may be important at different moments in time. As the angles themselves may not be so easy to interpret, we also plot the elements of the first eigenvector in Figure 6. As all elements of this vector turn out to be positive at all times, we plot the squared elements in a stacked way, given the unit length of the eigenvector itself. The firms with the largest systematic risk (SRISK) are plotted at the bottom. There is clear variation over time in the relative importance of the different stocks in the first eigenvector. Most of the globally active international banks are at the bottom given their large SRISK, whereas the top five positions are taken by lenders that are more locally active and concentrated on retail. This also

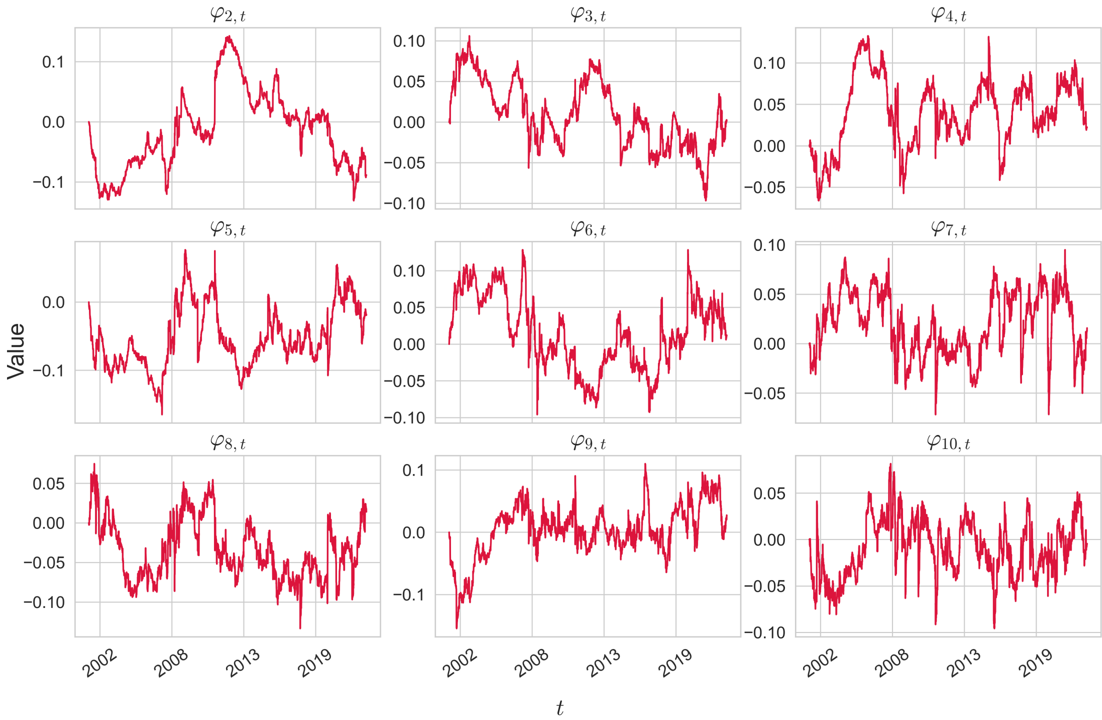

Figure 5: Angle dynamics

This figure shows the time dynamics in the rotation parameters $\varphi_{2,t}, \ldots, \varphi_{10,t}$. The time dynamics are shown for the full-sample period from January 2002 to February 2023.

shows in the elements of the *second* eigenfactor (see Appendix C), where the former and latter financial institutions have loadings with opposite signs.

Looking at Figure 6, we see that the combined squared weights of the first five banks range from as low as 40% in the early 2000s and after 2020, to as high as 65% in 2012. The weight of individual banks may also vary substantially over time, as can be seen, for example, from the width of the band for BAC, which ranges from about 0.05 in 2003 to 0.2 in 2012. We also observe clear responses in the weights around several major financial market events during the sample period. For instance, the August 2007 BNP freeze of payments, marking the start of the financial crisis, substantially increases the proportion of the less systemically important banks (plotted at the top) in the first eigenvector, as the entire financial sector went through turbulent times. The fall of Lehman, on the other hand, seems to have a much less persistent effect. The US downgrade by S&P in August 2011,

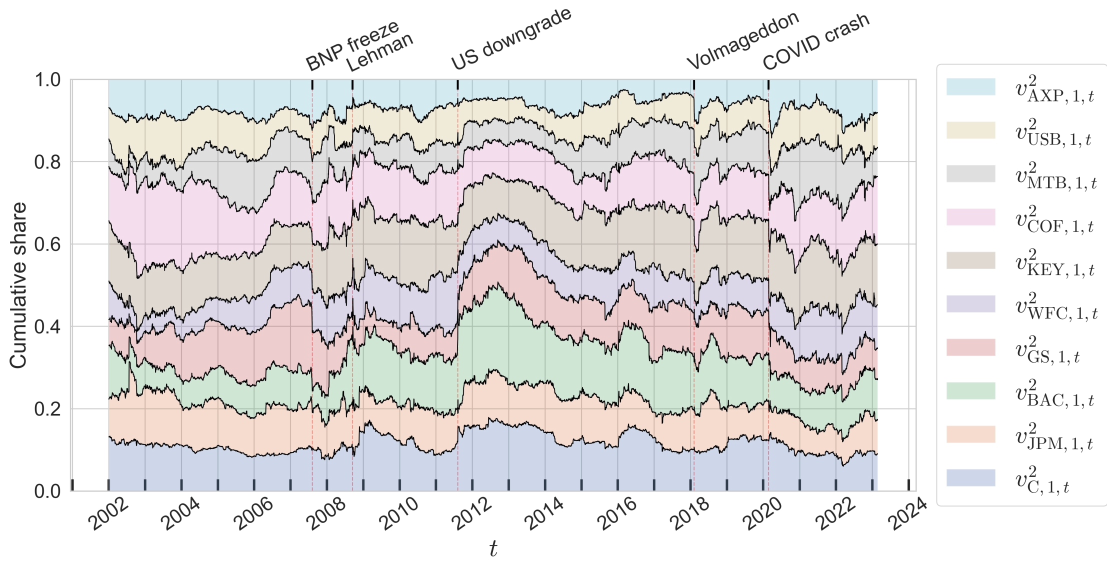

Figure 6: Dynamics of eigenvector composition

This figure shows the time series of all squared elements of the first eigenvector of the conditional covariance matrix. The full-sample analysis spans the period January 2002 to January 2023. The assets are ordered by mean SRISK over the sample period. The asset with the highest SRISK is presented at the bottom.

the doubling of the CBOE VIX index (Volmageddon) in February 2018, and the COVID-19 lockdowns and FED responses in early 2020, all have visible effects on the direction of the eigenvector. Not all these effects are the same, though. Whereas the US downgrade seems to concentrate the eigenvector more towards the systemically most important and internationally most active banks, the COVID event instead pushes the eigenvector to a more equally weighted scenario. The current DSR model helps uncover these effects and facilitates further analysis.

Finally, we compare the new DSR model with the benchmark models from Section 5.3 in an out-of-sample analysis. Table 3 presents the results. Panel A of Table 3 clearly shows that the DSR outperforms its competitors in terms of log-scores, Frobenius norm, and QLIKE: most of the DM statistics are statistically significant in favor of the DSR model. Panel B of the same table shows that also in economic terms, using the ex-post variance of the Global Minimum Variance (GMV) portfolio returns, the DSR model performs well. Using the realized portfolio variance measures ($\sqrt{rc_{pf}}$), the DSR is clearly significantly better. These economic

Table 3: Out-of-Sample performance

This table shows the results of the predictive application for 10 financial US equities. The performance is shown for the Gaussian and Student's t-distribution in the left and right panels, respectively. A comparison is made between the proposed DSR model and the benchmark models: the λ-GARCH model of Hetland et al. (2023) and the BEKK model of Engle and Kroner (1995). The results are based on predictions over a 16-year out-of-sample period of data spanning from January 2007 to February 2023, with re-estimation of the model every 20 days. The rows represent the performance measures and are divided into two panels: Panel A contains matrix forecasting measures, and Panel B contains an economic value comparison. Panel A shows the out-of-sample log-scores, the average Frobenius-normed difference between the forecasted covariance matrix and the observed (estimated) realized covariance matrix, and the mean QLIKE. Panel B shows the standard deviation of the Global Minimum Variance Portfolio (GMVP) returns, the mean realized GMV portfolio volatility, the mean portfolio turnover, and the mean portfolio concentration. The Diebold-Mariano (DM) test statistic and p-values are also shown for each of the measures and benchmark models in comparison to the DSR model. Stars indicate statistical significance of the DM test against the DSR model, applied to Log-score, Frob, QLIKE, StDev r_{pf} (tested on $\{r_{pf}^2\}$), and $\{\sqrt{rc_{pf}}\}$: $^{*}p < 0.1$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

	Gaussian			Student's t		
	DSR	λ-GARCH	BEKK	DSR	λ-GARCH	BEKK
Panel A: Likelihood						
Log-score	-14.588	-14.715^{***}	-14.676^{*}	-14.028	-14.190^{***}	-14.067^{**}
Frob	23.735	24.043^{***}	25.926^{***}	23.459	23.541^{*}	23.818
QLIKE	14.323	14.312	16.214^{***}	16.692	16.779^{***}	22.581^{***}
Panel B: Economic Value Comparison						
StDev r_{pf}	1.602	1.680^{**}	1.544	1.554	1.656^{***}	1.524
$\sqrt{\overline{rc}_{pf}}$	1.650	1.706^{***}	1.733^{***}	1.623	1.679^{***}	1.709^{***}
$\overline{\text{TO}}_t$	0.239	0.152	0.176	0.219	0.134	0.192
$\overline{\text{CO}}_t$	0.806	0.805	0.890	0.818	0.802	0.894

improvements in realized variances are obtained at about the same or even a lower level of GMV portfolio concentration (CO) and a somewhat higher portfolio turnover. We conclude that also out-of-sample, the new DSR gives good and robust out-of-sample results compared to its competitors. Only if daily ex-post returns are used rather than realized variances, the BEKK model is slightly better, but the difference is insignificant in this case.

6 Conclusion

We introduced the new Dynamic Spectral Rotation (DSR) model for dynamic covariance matrices by allowing both eigenvalues and the principal eigenvector to vary over time through score-driven dynamics. We developed a full asymptotic framework for this model. In particular, we obtained a new identification result based on a dynamic rather than static identification strategy, which explains why large variations in the rotation angles of the eigenvectors, even outside the $[-\pi, \pi]$ range, can still be recovered from the data. The asymptotic results carried over accurately to finite samples, as shown in our controlled simulation study.

In an empirical application to 10 US financial stocks, the new DSR model delivered a favorable out-of-sample performance relative to two benchmark models, illustrating that allowing for time variation in the leading eigenvector can be empirically relevant, both statistically and economically. The filtered time-varying first eigenvector clearly aligned with some of the major events in financial markets over this period.

Looking forward, we mention two directions for future research. First, although our empirical results show that a generalization of the current model to a setting with multiple time-varying eigenvectors is easily implementable, it is also interesting to see whether the new theoretical identification results can be generalized to such a setting. For this, one can try to further exploit the column structure of the products of Givens rotation matrices. Second, our approach to model eigenvectors as dynamic objects might also be relevant in other contexts where eigenvalues and eigenvectors play a dominant role, and it is interesting to explore whether the current modeling strategy could have applications there as well. We leave these extensions for now for a future paper.

References

Artemova, M. (2025). An order-invariant score-driven dynamic factor model. *Journal of Econometrics 251*, 106073.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2009). Realized kernels in practice: Trades and quotes. *Econometrics Journal 12*(3), C1–C32.

Beutner, E., Y. Lin, and A. Lucas (2026). Consistency, distributional convergence, and optimality of time-varying parameters in score-driven models. *Journal of Econometrics 255*, 106218.

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

Blasques, F., J. van Brummelen, S. J. Koopman, and A. Lucas (2022). Maximum likelihood estimation for score-driven models. *Journal of Econometrics 227*(2), 325–346.

Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model. *Review of Economics and Statistics*, 498–505.

Bollerslev, T., A. J. Patton, and R. Quaedvlieg (2018). Modeling and forecasting (un) reliable realized covariances for more reliable financial decisions. *Journal of Econometrics 207*(1), 71–91.

Brownlees, C. T. and G. M. Gallo (2006). Financial econometric analysis at ultra-high frequency: Data handling concerns. *Computational Statistics & Data Analysis 51*(4), 2232–2245.

Chiriac, R. and V. Voev (2011). Modelling and forecasting multivariate realized volatility. *Journal of Applied Econometrics 26*(6), 922–947.

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*(4), 552–563.

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

Creal, D., S. J. Koopman, A. Lucas, and M. Zamojski (2024). Observation-driven filtering of time-varying parameters using moment conditions. *Journal of Econometrics 238*(2), 105635.

DeMiguel, V., L. Garlappi, and R. Uppal (2009). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *Review of Financial Studies 22*(5), 1915–1953.

Diebold, F. X. and R. S. Mariano (2002). Comparing predictive accuracy. *Journal of Business & Economic Statistics 20*(1), 134–144.

Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics 20*(3), 339–350.

Engle, R. and K. Kroner (1995). Multivariate simultaneous generalized ARCH. *Econometric Theory 11*(1), 122–150.

Goetzmann, W. N. and A. Kumar (2008). Equity portfolio diversification. *Review of Finance 12*(3), 433–463.

Gorgi, P., C. Lauria, and A. Luati (2024). On the optimality of score-driven models. *Biometrika 111*(3), 865–880.

Gubbels, K. B. and A. Lucas (2026). Spectral dynamics and regularization for high-dimensional copulas. *arXiv preprint arXiv:2601.13281*.

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.

Hautsch, N., L. M. Kyj, and P. Malec (2015). Do high-frequency data improve high-dimensional portfolio allocations? *Journal of Applied Econometrics 30*(2), 263–290.

Hetland, S. (2020). Spectral targeting estimation of λ-GARCH models. *arXiv preprint arXiv:2007.02588*.

Hetland, S., R. S. Pedersen, and A. Rahbek (2023). Dynamic conditional eigenvalue GARCH. *Journal of Econometrics 237*(2), 105175.

Ledoit, O. and M. Wolf (2022). Quadratic shrinkage for large covariance matrices. *Bernoulli 28*, 1519–1547.

Lin, Y. and A. Lucas (2025). Testing for the absence of score-driven parameter dynamics. Technical Report TI 2025-063/III, Tinbergen Institute Discussion Paper.

Pinheiro, J. C. and D. M. Bates (1996). Unconstrained parametrizations for variance-covariance matrices. *Statistics and Computing 6*, 289–296.

Shephard, N. and K. Sheppard (2010). Realising the future: forecasting with high-frequency-based volatility (heavy) models. *Journal of Applied Econometrics 25*(2), 197–231.

Van der Weide, R. (2002). GO-GARCH: a multivariate generalized orthogonal GARCH model. *Journal of Applied Econometrics 17*(5), 549–564.

Supplementary Materials for

"Joint Eigenvector and Eigenvalue Dynamics with an Application to Time-Varying Covariance Matrices"

This supplement includes all technical proofs and additional empirical results.

Contents

A Proofs of the main results

Notation: For a complete separable metric space denoted by (E, d_E), we define the Lipschitz coefficient ρ associated with a random map $h: E \to E$ as:

$$\rho(h) = \sup_{x,y \in E,\, x \neq y} \left\{ \frac{d_E\big(h(x), h(y)\big)}{d_E(x, y)} \right\}. \tag{A.1}$$

If h is measurable, then so is $\rho(h)$ (Bougerol, 1993, p. 955). This definition will be used repeatedly in the subsequent sections, with the specific space E being identified in each instance. We use the notation for matrix derivatives suggested by Magnus and Neudecker (2019):

$$\mathrm{D}_{\boldsymbol{X}} \boldsymbol{F}(\boldsymbol{X}) = \frac{\partial \operatorname{vec}(\boldsymbol{F}(\boldsymbol{X}))}{\partial \operatorname{vec}(\boldsymbol{X})^\top}. \tag{A.2}$$

Finally, the symbol C denotes a generic positive constant, which may differ from one occurrence to another.

Proof of Proposition 1. Note that, by definition, for $\boldsymbol{\varphi} = (\varphi_2, \dots, \varphi_k)^\top$, one has $\mathrm{d}\boldsymbol{S}(\boldsymbol{\varphi}) = \sum_{i=2}^{k} \mathscr{D}_i \boldsymbol{S}(\boldsymbol{\varphi})\, \mathrm{d}\varphi_i$, where $\mathscr{D}_i \boldsymbol{S}(\boldsymbol{\varphi})$ is provided in (2.6). Recall that $\boldsymbol{V}_t = \bar{\boldsymbol{V}} \boldsymbol{S}_t$. Using the fact that $\mathrm{d} \log \det(\boldsymbol{X}) = \operatorname{tr}(\boldsymbol{X}^{-1} \mathrm{d}\boldsymbol{X})$ and $\mathrm{d}\boldsymbol{X}^{-1} = -\boldsymbol{X}^{-1}(\mathrm{d}\boldsymbol{X})\boldsymbol{X}^{-1}$ for any matrix $\boldsymbol{X}$, if $\boldsymbol{y}_t \mid \mathscr{F}_{t-1} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_t)$, we obtain

$$\begin{aligned}
\mathrm{d}\ell_t(\boldsymbol{f}_t, \boldsymbol{\nu}) &= -\frac{1}{2} \operatorname{tr}\left(\boldsymbol{\Lambda}_t^{-1} \mathrm{d}\boldsymbol{\Lambda}_t\right) - \frac{1}{2} \boldsymbol{y}_t^\top \boldsymbol{V}_t \big(\mathrm{d}\boldsymbol{\Lambda}_t^{-1}\big) \boldsymbol{V}_t^\top \boldsymbol{y}_t - \boldsymbol{y}_t^\top (\mathrm{d}\boldsymbol{V}_t) \boldsymbol{\Lambda}_t^{-1} \boldsymbol{V}_t^\top \boldsymbol{y}_t \\
&= \frac{1}{2} \sum_{i=1}^{k} \lambda_{i,t}^{-2} \Big[\big(\boldsymbol{V}_t^\top \boldsymbol{y}_t\big)_i^2 - \lambda_{i,t} \Big] \mathrm{d}\lambda_{i,t} - \sum_{i=2}^{k} \boldsymbol{y}_t^\top \bar{\boldsymbol{V}} \big[\mathscr{D}_i \boldsymbol{S}(\boldsymbol{\varphi}_t)\big] \boldsymbol{\Lambda}_t^{-1} \boldsymbol{V}_t^\top \boldsymbol{y}_t \, \mathrm{d}\varphi_{i,t}.
\end{aligned}$$

Similarly, if $\boldsymbol{y}_t \mid \mathscr{F}_{t-1} \sim t_\nu\big(\boldsymbol{0}, \frac{\nu-2}{\nu} \boldsymbol{\Sigma}_t\big)$, by (2.5), we have

$$\begin{aligned}
\mathrm{d}\ell_t(\boldsymbol{f}_t, \boldsymbol{\nu}) &= -\frac{1}{2} \operatorname{tr}\left(\boldsymbol{\Lambda}_t^{-1} \mathrm{d}\boldsymbol{\Lambda}_t\right) - \frac{1}{2} w_t \Big(2\boldsymbol{y}_t^\top (\mathrm{d}\boldsymbol{V}_t) \boldsymbol{\Lambda}_t^{-1} \boldsymbol{V}_t^\top \boldsymbol{y}_t + \boldsymbol{y}_t^\top \boldsymbol{V}_t \big(\mathrm{d}\boldsymbol{\Lambda}_t^{-1}\big) \boldsymbol{V}_t^\top \boldsymbol{y}_t \Big) \\
&= -\frac{1}{2} \Big(\operatorname{tr}\left(\boldsymbol{\Lambda}_t^{-1} \mathrm{d}\boldsymbol{\Lambda}_t\right) + w_t\, \boldsymbol{y}_t^\top \boldsymbol{V}_t \big(\mathrm{d}\boldsymbol{\Lambda}_t^{-1}\big) \boldsymbol{V}_t^\top \boldsymbol{y}_t \Big) - w_t\, \boldsymbol{y}_t^\top (\mathrm{d}\boldsymbol{V}_t) \boldsymbol{\Lambda}_t^{-1} \boldsymbol{V}_t^\top \boldsymbol{y}_t
\end{aligned}$$

$$= \frac{1}{2}\sum_{i=1}^{k}\lambda_{i,t}^{-2}\Big[w_t\big(\boldsymbol{V}_t^\top\boldsymbol{y}_t\big)_i^2 - \lambda_{i,t}\Big]\mathrm{d}\lambda_{i,t} - w_t\sum_{i=2}^{k}\boldsymbol{y}_t^\top\bar{\boldsymbol{V}}\big[\mathscr{D}_i\boldsymbol{S}(\boldsymbol{\varphi}_t)\big]\boldsymbol{\Lambda}_t^{-1}\boldsymbol{V}_t^\top\boldsymbol{y}_t\,\mathrm{d}\varphi_{i,t},$$

which in turn yields the results presented in Proposition 1. □

In the proofs below, we repeatedly use the following inequalities (see, e.g., Lin and Lucas, 2025a, (B.3)–(B.4)): for any matrices $\boldsymbol{X}_t$, $t = 1, \ldots, K$, where $K \in \mathbb{Z}^+$ and the dimensions are compatible,

$$\log^+\left\|\prod_{t=1}^{K}\boldsymbol{X}_t\right\| \leqslant \sum_{t=1}^{K}\log^+\|\boldsymbol{X}_t\|, \qquad \log^+\left\|\sum_{t=1}^{K}\boldsymbol{X}_t\right\| \leqslant \log(K) + \sum_{t=1}^{K}\log^+\|\boldsymbol{X}_t\|. \tag{A.3}$$

Proof of Proposition 2. For Proposition 2(i), we view $\widehat{\boldsymbol{\lambda}}_t^\varepsilon$ for $t \in \mathbb{Z}^+$ as an element of the complete and separable metric space $(E, d_E) = (\mathbb{F}_\lambda, \|\cdot\|)$. Write the SRE in (3.1) as $\widehat{\boldsymbol{\lambda}}_{t+1}^\varepsilon = \boldsymbol{\psi}_t^{\varepsilon,\lambda}(\widehat{\boldsymbol{\lambda}}_t^\varepsilon)$, where $\boldsymbol{\psi}_t^{\varepsilon,\lambda}(\boldsymbol{\lambda}) = (\boldsymbol{I}_k - \boldsymbol{B}_\lambda)\boldsymbol{\omega}_\lambda + \boldsymbol{\Phi}_t\,\boldsymbol{\lambda}$, where $\boldsymbol{\Phi}_t = (\boldsymbol{B}_\lambda - \boldsymbol{A}_\lambda) + \boldsymbol{A}_\lambda\,\mathrm{diag}(\boldsymbol{\varepsilon}_t^{\odot 2})$ is defined below (3.1). Since $\boldsymbol{\varepsilon}_t$ for $t \in \mathbb{Z}$ are i.i.d., the sequence $\{\boldsymbol{\psi}_t^{\varepsilon,\lambda}, t \in \mathbb{Z}\}$ is SE for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ (see, e.g., White, 2001, Theorem 3.35). The convergence of $\big\{\widehat{\boldsymbol{\lambda}}_t^\varepsilon, t \in \mathbb{Z}^+\big\}$ then can be obtained from Theorem 3.1 in Bougerol (1993) by verifying the following conditions. For $\boldsymbol{\theta} \in \boldsymbol{\Theta}$,

(a) $\mathbb{E}\big[\log^+\big\|\boldsymbol{\psi}_1^{\varepsilon,\lambda}(\widehat{\boldsymbol{\lambda}}_1^\varepsilon) - \widehat{\boldsymbol{\lambda}}_1^\varepsilon\big\|\big] < \infty$ for some nonrandom $\widehat{\boldsymbol{\lambda}}_1^\varepsilon \in \mathbb{F}_\lambda$, where ρ is defined in (A.1);

(b) $\mathbb{E}\big[\log^+\rho\big(\boldsymbol{\psi}_t^{\varepsilon,\lambda}\big)\big] < \infty$;

(c) $\mathbb{E}\big[\log\rho\big(\boldsymbol{\psi}_t^{\varepsilon,\lambda,(r)}\big)\big] < 0$, where $\boldsymbol{\psi}_t^{\varepsilon,\lambda,(r)}(\cdot) = \boldsymbol{\psi}_t^{\varepsilon,\lambda}(\cdot) \circ \boldsymbol{\psi}_{t-1}^{\varepsilon,\lambda}(\cdot) \circ \ldots \circ \boldsymbol{\psi}_{t-r+1}^{\varepsilon,\lambda}(\cdot)$ is the r-fold convolution of $\boldsymbol{\psi}_t^{\varepsilon,\lambda}$ for some integer $r \geqslant 1$.

Note that $\mathbb{E}\big[\log^+\big\|\boldsymbol{\psi}_1^{\varepsilon,\lambda}(\widehat{\boldsymbol{\lambda}}_1^\varepsilon) - \widehat{\boldsymbol{\lambda}}_1^\varepsilon\big\|\big] < \infty$ in condition ***(a)*** immediately follows from (A.3), and the condition that $\mathbb{E}\,\|\boldsymbol{\varepsilon}_t\|^{2\eta} < \infty$ for some $\eta > 0$. For the condition ***(b)***, by applying a mean value theorem for vector-valued functions (e.g., Rudin, 1976, Theorem 9.19), we obtain $\rho\big(\boldsymbol{\psi}_t^{\varepsilon,\lambda}\big) \leqslant \sup_{\boldsymbol{\lambda}\in\mathbb{F}_\lambda}\big\|\partial\boldsymbol{\psi}_t^{\varepsilon,\lambda}(\boldsymbol{\lambda})/\partial\boldsymbol{\lambda}^\top\big\| = \big\|\boldsymbol{\Phi}_t\big\|$. It follows immediately that $\mathbb{E}\big[\log^+\rho\big(\boldsymbol{\psi}_t^{\varepsilon,\lambda}\big)\big] < \infty$ under the condition that $\mathbb{E}\|\boldsymbol{\varepsilon}_t\|^{2\eta} <$

∞ for some $\eta > 0$. Finally, condition ***(c)*** is implied by the assumption that $\mathbb{E}\big[\log\big(\|\prod_{t=1}^{r}\boldsymbol{\Phi}_t\|\big)\big] < 0$ for some integer $r \geqslant 1$.

To prove Proposition 2(ii), note that $\mathbb{E}\big(\|\prod_{t=1}^{r}\boldsymbol{\Phi}_t\|^{\eta}\big) < 1$ implies $\mathbb{E}\big[\log\big(\|\prod_{t=1}^{r}\boldsymbol{\Phi}_t\|\big)\big] < 0$ by Jensen's inequality. Therefore, the results in Part (i) continue to hold. Before continuing, we first show that

$$\mathbb{E}\big(\|\boldsymbol{\lambda}_t^{\varepsilon}\|^{\eta}\big) < \infty. \tag{A.4}$$

Note that $\boldsymbol{\lambda}_t^{\varepsilon}$ admits an almost sure representation as a measurable function of $\{\boldsymbol{\varepsilon}_{t-1}, \boldsymbol{\varepsilon}_{t-2}, \ldots\}$, i.e., $\boldsymbol{\lambda}_t^{\varepsilon} = \sum_{j=0}^{\infty}\big(\prod_{\ell=1}^{j}\boldsymbol{\Phi}_{t-\ell}\big)(\boldsymbol{I}_k - \boldsymbol{B}_\lambda)\boldsymbol{\omega}_\lambda$ with $\prod_{\ell=1}^{0}\cdot \equiv \boldsymbol{I}_k$. If $\eta < 1$, we have $\mathbb{E}\big(\|\boldsymbol{\lambda}_t^{\varepsilon}\|^{\eta}\big) \leqslant \|(\boldsymbol{I}_k - \boldsymbol{B}_\lambda)\boldsymbol{\omega}_\lambda\|^{\eta}\sum_{j=0}^{\infty}\mathbb{E}\|\prod_{\ell=1}^{j}\boldsymbol{\Phi}_{t-\ell}\|^{\eta}$ by the c_r-inequality. Following the same argument as that between Eqs. (3.17) and (3.18) in Straumann and Mikosch (2006), we obtain $\mathbb{E}\|\prod_{\ell=1}^{j}\boldsymbol{\Phi}_{t-\ell}\|^{\eta} \leqslant C_{\eta,r}\,\delta^{[j/r]}$ for some $\delta \in (0,1)$ and all integers $j \geqslant 0$, where $C_{\eta,r} = \max\big\{1, \big(\mathbb{E}\,\|\boldsymbol{\Phi}_t\|^{\eta}\big)^{r}\big\} < \infty$ with $r \geqslant 1$. Therefore, $\mathbb{E}\big(\|\boldsymbol{\lambda}_t^{\varepsilon}\|^{\eta}\big) \leqslant \|(\boldsymbol{I}_k - \boldsymbol{B}_\lambda)\boldsymbol{\omega}_\lambda\|^{\eta}\,C_{\eta,r}\sum_{j=0}^{\infty}\delta^{[j/r]} < \infty$. For $\eta \geqslant 1$, applying Minkowski's inequality yields $\mathbb{E}\big(\|\boldsymbol{\lambda}_t^{\varepsilon}\|^{\eta}\big) \leqslant \|(\boldsymbol{I}_k - \boldsymbol{B}_\lambda)\boldsymbol{\omega}_\lambda\|^{\eta}\,C_{\eta,r}^{\eta}\big[\sum_{j=0}^{\infty}\big(\delta^{1/\eta}\big)^{[j/r]}\big]^{\eta} < \infty$. Hence, (A.4) holds.

To continue, we first show that the joint process $\big\{\big(\boldsymbol{f}_t^{\varepsilon\top}, \boldsymbol{\varepsilon}_t^{\top}\big)^{\top}, t \in \mathbb{Z}\big\}$ is SE. If this holds, then the SE property of $\{\boldsymbol{y}_t^{\varepsilon}, t \in \mathbb{Z}\}$ follows immediately, since $\boldsymbol{y}_t^{\varepsilon}$ is a continuous, and hence measurable, function of $\big(\boldsymbol{f}_t^{\varepsilon\top}, \boldsymbol{\varepsilon}_t^{\top}\big)^{\top}$. To this end, we show that the eigenvector dynamics in (3.2), viewed as an SRE with input $\{(\boldsymbol{\lambda}_t^{\varepsilon\top}, \boldsymbol{\varepsilon}_t^{\top})^{\top}, t \in \mathbb{Z}\}$, admits a unique SE solution $\boldsymbol{\varphi}_t^{\varepsilon}$ to which any initialized sequence converges e.a.s. If this holds, then $\boldsymbol{\varphi}_t^{\varepsilon}$ is measurable with respect to $\big\{(\boldsymbol{\lambda}_s^{\varepsilon\top}, \boldsymbol{\varepsilon}_s^{\top})^{\top}, s \leqslant t-1\big\}$, and hence is also ultimately measurable with respect to $\{\boldsymbol{\varepsilon}_{t-1}, \boldsymbol{\varepsilon}_{t-2}, \ldots\}$. Therefore, $\big\{\big(\boldsymbol{f}_t^{\varepsilon\top}, \boldsymbol{\varepsilon}_t^{\top}\big)^{\top}, t \in \mathbb{Z}\big\}$ is a measurable function of $\{\boldsymbol{\varepsilon}_t, \boldsymbol{\varepsilon}_{t-1}, \boldsymbol{\varepsilon}_{t-2}, \ldots\}$, and is thus SE (White, 2001, Theorem 3.35). For $t \geqslant 2$, let $\widehat{\boldsymbol{\varphi}}_t^{\varepsilon}$ be an initialized sequence with nonrandom initial value $\widehat{\boldsymbol{\varphi}}_1^{\varepsilon} \in \mathbb{R}^{k-1}$, and suppose that it follows the SRE in (3.2):

$$\widehat{\boldsymbol{\varphi}}_{t+1}^{\varepsilon} = \boldsymbol{\psi}_t^{\varepsilon,\varphi}(\widehat{\boldsymbol{\varphi}}_t^{\varepsilon}) = (\boldsymbol{I}_{k-1} - \boldsymbol{B}_{\varphi})\boldsymbol{\omega}_{\varphi} + \boldsymbol{B}_{\varphi}\,\widehat{\boldsymbol{\varphi}}_t^{\varepsilon} + \boldsymbol{A}_{\varphi}\,\boldsymbol{s}_t^{\varepsilon,\varphi}(\hat{\boldsymbol{f}}_t^{\varepsilon},\boldsymbol{\theta}), \quad \hat{\boldsymbol{f}}_t^{\varepsilon} = \begin{pmatrix} \boldsymbol{\lambda}_t^{\varepsilon} \\ \widehat{\boldsymbol{\varphi}}_t^{\varepsilon} \end{pmatrix}, \tag{A.5}$$

where $\{\boldsymbol{\psi}_t^{\varepsilon,\varphi},\, t \in \mathbb{Z}\}$ is SE for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ because $\{(\boldsymbol{\lambda}_s^{\varepsilon\top}, \boldsymbol{\varepsilon}_t^{\top})^{\top}, t \in \mathbb{Z}\}$ is SE. Similar to the proof of Part (i), $\mathbb{E}\big[\log^+ \|\boldsymbol{\psi}_1^{\varepsilon,\varphi}(\widehat{\boldsymbol{\varphi}}_1^{\varepsilon}) - \widehat{\boldsymbol{\varphi}}_1^{\varepsilon}\|\big] < \infty$ follows from the moment condition on $\boldsymbol{\varepsilon}_t$ in Proposition 2, the lower bound of $\mathbb{F}_{\lambda}$ as required in Part (ii), and (A.4) above. Moreover, by a mean value theorem for vector-valued functions (e.g., Rudin, 1976, Theorem 9.19) and the contraction condition imposed in Part (ii), it is not hard to obtain the contraction condition $\mathbb{E}\big[\log \rho\big(\boldsymbol{\psi}_t^{\varepsilon,\varphi}\big)\big] < 0$; see also the proof of Proposition 3 for more details. It follows that $\{\widehat{\boldsymbol{\varphi}}_t^{\varepsilon},\, t \in \mathbb{Z}^+\}$ converges e.a.s. to the unique SE solution $\{\boldsymbol{\varphi}_t^{\varepsilon},\, t \in \mathbb{Z}\}$ defined by (3.2).

Finally, we establish the existence of moments of $\boldsymbol{y}_t^{\varepsilon}$. Note that $\|\boldsymbol{S}(\boldsymbol{\varphi})\| = 1$ for any $\boldsymbol{\varphi} \in \mathbb{R}^{k-1}$. Since $\boldsymbol{\lambda}_t^{\varepsilon}$ is measurable with respect to $\{\boldsymbol{\varepsilon}_{t-1}, \boldsymbol{\varepsilon}_{t-2}, \ldots\}$ and is therefore independent of $\boldsymbol{\varepsilon}_t$, we have $\mathbb{E}\,\|\boldsymbol{y}_t^{\varepsilon}\|^{2\eta} \leqslant \mathbb{E}\big(\|\boldsymbol{\lambda}_t^{\varepsilon}\|_{\infty}^{\eta}\,\|\boldsymbol{\varepsilon}_t\|^{2\eta}\big) = \mathbb{E}\big(\|\boldsymbol{\lambda}_t^{\varepsilon}\|_{\infty}^{\eta}\big)\mathbb{E}\big(\|\boldsymbol{\varepsilon}_t\|^{2\eta}\big)$. Because $\mathbb{E}\,\|\boldsymbol{\varepsilon}_t\|^{2\eta} < \infty$ and $\|\boldsymbol{\lambda}_t^{\varepsilon}\|_{\infty} \leqslant \|\boldsymbol{\lambda}_t^{\varepsilon}\|$, $\mathbb{E}\,\|\boldsymbol{y}_t^{\varepsilon}\|^{2\eta} < \infty$ follows immediately from (A.4). This completes the proof. $\square$

Additional notation is required to establish the invertibility of the filter. Let $\mathcal{C}^0\big(\boldsymbol{\Theta}, \mathbb{F}\big)$ denote the space of all continuous $\mathbb{F}$-valued functions on $\boldsymbol{\Theta}$, equipped with the supremum norm $\|\cdot\|_{\boldsymbol{\Theta}}$, defined by $\|\boldsymbol{f}\|_{\boldsymbol{\Theta}} = \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\boldsymbol{f}(\boldsymbol{\theta})\|$ for $\boldsymbol{f} \in \mathcal{C}^0\big(\boldsymbol{\Theta}, \mathbb{F}\big)$. Moreover, for any measurable map $\boldsymbol{f} : \boldsymbol{\Theta} \to \mathbb{R}^{m+n}$ with block decomposition $\boldsymbol{f}(\boldsymbol{\theta}) = \big(\boldsymbol{u}(\boldsymbol{\theta})^{\top}, \boldsymbol{v}(\boldsymbol{\theta})^{\top}\big)^{\top}$, where $\boldsymbol{u}(\boldsymbol{\theta}) \in \mathbb{R}^m$ and $\boldsymbol{v}(\boldsymbol{\theta}) \in \mathbb{R}^n$, we define the weighted supremum norm $\|\cdot\|_{\alpha}$ for $\alpha > 0$ as

$$\|\boldsymbol{f}\|_{\boldsymbol{\Theta},\alpha} := \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\boldsymbol{f}(\boldsymbol{\theta})\|_{\alpha}, \qquad \|\boldsymbol{f}(\boldsymbol{\theta})\|_{\alpha} := \|\boldsymbol{u}(\boldsymbol{\theta})\| + \alpha\,\|\boldsymbol{v}(\boldsymbol{\theta})\|. \tag{A.6}$$

By the norm equivalence established in Lemma B.1 and by the compactness of $\boldsymbol{\Theta}$ required in Assumption 1(i), it follows that $\big(\mathcal{C}^0\big(\boldsymbol{\Theta}, \mathbb{F}\big), \|\cdot\|_{\boldsymbol{\Theta},\alpha}\big)$ is complete and separable, and hence a Polish space; see Kechris (2012, Theorem 4.19). Finally, for any matrix $\boldsymbol{A}$, let $\|\boldsymbol{A}\|_\alpha := \sup_{\boldsymbol{x}\neq\boldsymbol{0}} \|\boldsymbol{A}\boldsymbol{x}\|_\alpha / \|\boldsymbol{x}\|_\alpha$ be the induced operator norm.

Proof of Proposition 3. For $\alpha > 0$, we regard the sequence $\{\hat{\boldsymbol{f}}_t(\cdot),\, t \in \mathbb{Z}^+\}$, initialized at some nonrandom $\hat{\boldsymbol{f}}_1(\cdot) \equiv \hat{\boldsymbol{f}}_1 \in \mathbb{F}$, as a sequence of random elements taking values in the Polish space $(E, d_E) = \big(\mathcal{C}^0(\boldsymbol{\Theta}, \mathbb{F}), \|\cdot\|_{\boldsymbol{\Theta},\alpha}\big)$; see Eq. (A.1). Write $\boldsymbol{f} = (\boldsymbol{\lambda}^\top, \boldsymbol{\varphi}^\top)^\top$, where the first component corresponds to the eigenvalues and the second to the parameters determining the eigenvectors, where $\boldsymbol{\varphi} = (\varphi_2, \ldots, \varphi_k)^\top$. For convenience, let $\bar{\boldsymbol{y}}_t = \bar{\boldsymbol{V}}^\top \boldsymbol{y}_t$ and $\boldsymbol{s}_t(\boldsymbol{f}) = \begin{pmatrix} \boldsymbol{s}_t^\lambda(\boldsymbol{f}) \\ \boldsymbol{s}_t^\varphi(\boldsymbol{f}) \end{pmatrix}$, where

$$\boldsymbol{s}_t^\lambda(\boldsymbol{f}) = \big(\boldsymbol{S}(\boldsymbol{\varphi})^\top \bar{\boldsymbol{y}}_t\big)^{\odot 2} - \boldsymbol{\lambda}, \quad \boldsymbol{s}_t^\varphi(\boldsymbol{f}) = -\begin{pmatrix} \bar{\boldsymbol{y}}_t^\top [\mathscr{D}_2 \boldsymbol{S}(\boldsymbol{\varphi})] \boldsymbol{\Lambda}^{-1} [\boldsymbol{S}(\boldsymbol{\varphi})]^\top \bar{\boldsymbol{y}}_t \\ \vdots \\ \bar{\boldsymbol{y}}_t^\top [\mathscr{D}_k \boldsymbol{S}(\boldsymbol{\varphi})] \boldsymbol{\Lambda}^{-1} [\boldsymbol{S}(\boldsymbol{\varphi})]^\top \bar{\boldsymbol{y}}_t \end{pmatrix}, \quad \boldsymbol{\Lambda} = \operatorname{diag}(\boldsymbol{\lambda}).$$

Then the Gaussian SRE $\boldsymbol{\psi}_t(\boldsymbol{f}, \boldsymbol{\theta})$ defined in (2.3) and Proposition 1 can be expressed as $\boldsymbol{\psi}_t(\boldsymbol{f}, \boldsymbol{\theta}) = (\boldsymbol{I}_{2k-1} - \boldsymbol{B})\boldsymbol{\omega} + \boldsymbol{B}\boldsymbol{f} + \boldsymbol{A}\,\boldsymbol{s}_t(\boldsymbol{f})$. Accordingly, $\hat{\boldsymbol{f}}_{t+1} = \boldsymbol{\mu}_t(\hat{\boldsymbol{f}}_t)$, where $\boldsymbol{\mu}_t : \big(\mathcal{C}^0(\boldsymbol{\Theta}, \mathbb{F}), \|\cdot\|_{\boldsymbol{\Theta},\alpha}\big) \to \big(\mathcal{C}^0(\boldsymbol{\Theta}, \mathbb{F}), \|\cdot\|_{\boldsymbol{\Theta},\alpha}\big)$ is a random mapping defined by $\big[\boldsymbol{\mu}_t(\boldsymbol{f})\big](\boldsymbol{\theta}) = \boldsymbol{\psi}_t(\boldsymbol{f}(\boldsymbol{\theta}), \boldsymbol{\theta})$. Since $\{\boldsymbol{y}_t,\, t \in \mathbb{Z}\}$ is SE, it follows that $\{\boldsymbol{\mu}_t,\, t \in \mathbb{Z}\}$ is also SE. As in Proposition 2, it suffices to verify the following conditions to apply Straumann and Mikosch (2006, Theorem 2.8):

(a) $\mathbb{E}\big[\log^+ \|\boldsymbol{\mu}_t(\hat{\boldsymbol{f}}_1) - \hat{\boldsymbol{f}}_1\|_{\boldsymbol{\Theta},\alpha}\big] < \infty$, where $\hat{\boldsymbol{f}}_1(\boldsymbol{\theta}) = \hat{\boldsymbol{f}}_1 \in \mathbb{F}$ for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$;

(b) $\mathbb{E}\big[\log \rho(\boldsymbol{\mu}_t)\big] < 0$.

To prove Condition ***(a)***, by Lemma B.1, (A.3), and the compactness of $\boldsymbol{\Theta}$, we obtain $\mathbb{E}\big[\log^+ \|\boldsymbol{\mu}_t(\hat{\boldsymbol{f}}_1) - \hat{\boldsymbol{f}}_1\|_{\boldsymbol{\Theta},\alpha}\big] < \infty$, provided that $\mathbb{E}\big[\log^+ \|\boldsymbol{s}_t^\lambda(\hat{\boldsymbol{f}}_1)\|\big] < \infty$ and $\mathbb{E}\big[\log^+ \|\boldsymbol{s}_t^\varphi(\hat{\boldsymbol{f}}_1)\|\big] < \infty$. Note that $\|\boldsymbol{S}(\boldsymbol{\varphi})\| = 1$ and $\|\mathscr{D}_i \boldsymbol{S}(\boldsymbol{\varphi})\| = 1$ for $i = 2, \ldots, k$. Moreover, by the definition $\bar{\boldsymbol{y}}_t = \bar{\boldsymbol{V}}^\top \boldsymbol{y}_t$, it follows that $\log^+ \big\|\big(\boldsymbol{S}(\boldsymbol{\varphi})^\top \bar{\boldsymbol{y}}_t\big)^{\odot 2}\big\| \leqslant \log^+ \big(\|\boldsymbol{S}(\boldsymbol{\varphi})\|^2 \|\bar{\boldsymbol{y}}_t\|^2\big) = 2\log^+ \|\boldsymbol{y}_t\|$. Since $\mathbb{E}\big[\log^+ \|\boldsymbol{y}_t\|\big] < \infty$, it follows immediately that $\mathbb{E}\big[\log^+ \|\boldsymbol{s}_t^\lambda(\hat{\boldsymbol{f}}_1)\|\big] < \infty$. Similarly, noting that $\log^+ \|\boldsymbol{s}_t^\varphi(\hat{\boldsymbol{f}}_1)\| \leqslant \log^+[\lambda_L^{-1}(k-1)] + 2\log^+ \|\boldsymbol{y}_t\|$, we have $\mathbb{E}\big[\log^+ \|\boldsymbol{s}_t^\varphi(\hat{\boldsymbol{f}}_1)\|\big] < \infty$.

To establish Condition ***(b)***, note that $\rho(\boldsymbol{\mu}_t) \leqslant \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sup_{\boldsymbol{f}\in\mathbb{F}} \left\| \partial \boldsymbol{\psi}_t(\boldsymbol{f}, \boldsymbol{\theta}) / \partial \boldsymbol{f}^\top \right\|_\alpha = \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sup_{\boldsymbol{f}\in\mathbb{F}} \left\| \boldsymbol{B} + \boldsymbol{A}\, \partial \boldsymbol{s}_t(\boldsymbol{f}) / \partial \boldsymbol{f}^\top \right\|_\alpha$ using a mean value theorem for vector-valued functions (Rudin, 1976, Theorem 9.19). Note that $\partial \boldsymbol{s}_t(\boldsymbol{f})/\partial \boldsymbol{f}^\top = \begin{pmatrix} \mathrm{D}_{\boldsymbol{\lambda}} \boldsymbol{s}_t^\lambda(\boldsymbol{f}) & \mathrm{D}_{\boldsymbol{\varphi}} \boldsymbol{s}_t^\lambda(\boldsymbol{f}) \\ \mathrm{D}_{\boldsymbol{\lambda}} \boldsymbol{s}_t^\varphi(\boldsymbol{f}) & \mathrm{D}_{\boldsymbol{\varphi}} \boldsymbol{s}_t^\varphi(\boldsymbol{f}) \end{pmatrix}$. By the block bound derived in Lemma B.2, we further obtain $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sup_{\boldsymbol{f}\in\mathbb{F}} \left\| \boldsymbol{B} + \boldsymbol{A}\, \partial \boldsymbol{s}_t(\boldsymbol{f}) / \partial \boldsymbol{f}^\top \right\|_\alpha \leqslant \max\left\{ Q_{1,t}(\alpha), Q_{2,t}(\alpha) \right\}$, where

$$Q_{1,t}(\alpha) = \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sup_{\boldsymbol{f}\in\mathbb{F}} \left\{ \left\| \boldsymbol{B}_\lambda + \boldsymbol{A}_\lambda\, \mathrm{D}_{\boldsymbol{\lambda}} \boldsymbol{s}_t^\lambda(\boldsymbol{f}) \right\| + \alpha \left\| \boldsymbol{A}_\varphi\, \mathrm{D}_{\boldsymbol{\lambda}} \boldsymbol{s}_t^\varphi(\boldsymbol{f}) \right\| \right\}, \tag{A.7}$$

$$Q_{2,t}(\alpha) = \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sup_{\boldsymbol{f}\in\mathbb{F}} \left\{ \left\| \boldsymbol{B}_\varphi + \boldsymbol{A}_\varphi\, \mathrm{D}_{\boldsymbol{\varphi}} \boldsymbol{s}_t^\varphi(\boldsymbol{f}) \right\| + \alpha^{-1} \left\| \boldsymbol{A}_\lambda\, \mathrm{D}_{\boldsymbol{\varphi}} \boldsymbol{s}_t^\lambda(\boldsymbol{f}) \right\| \right\}. \tag{A.8}$$

We therefore proceed to derive the components of $\partial \boldsymbol{s}_t(\boldsymbol{f})/\partial \boldsymbol{f}^\top$. Since $\mathrm{d}\boldsymbol{z}^{\odot 2} = 2\boldsymbol{z} \odot \mathrm{d}\boldsymbol{z}$, we have $\mathrm{d}\boldsymbol{s}_t^\lambda(\boldsymbol{f}) = 2\sum_{i=2}^k \Big[\big(\boldsymbol{S}(\boldsymbol{\varphi})^\top \bar{\boldsymbol{y}}_t \big) \odot \big([\mathscr{D}_i \boldsymbol{S}(\boldsymbol{\varphi})]^\top \bar{\boldsymbol{y}}_t \big) \Big] \mathrm{d}\varphi_i - \mathrm{d}\boldsymbol{\lambda}$. Therefore, $\mathrm{D}_{\boldsymbol{\lambda}} \boldsymbol{s}_t^\lambda(\boldsymbol{f}) = -\boldsymbol{I}_k$ and $\mathrm{D}_{\boldsymbol{\varphi}} \boldsymbol{s}_t^\lambda(\boldsymbol{f}) = 2\,\mathrm{diag}\left(\boldsymbol{S}(\boldsymbol{\varphi})^\top \bar{\boldsymbol{y}}_t \right) \Big([\mathscr{D}_2 \boldsymbol{S}(\boldsymbol{\varphi})]^\top \bar{\boldsymbol{y}}_t, \ldots, [\mathscr{D}_k \boldsymbol{S}(\boldsymbol{\varphi})]^\top \bar{\boldsymbol{y}}_t \Big)$. Furthermore, by $\mathrm{d}\boldsymbol{\Lambda}^m = m\boldsymbol{\Lambda}^{m-1}\mathrm{d}\boldsymbol{\Lambda}$ for any $m \in \mathbb{R}$, we obtain $-\mathrm{d}\bar{\boldsymbol{y}}_t^\top [\mathscr{D}_i \boldsymbol{S}(\boldsymbol{\varphi})] \boldsymbol{\Lambda}^{-1} [\boldsymbol{S}(\boldsymbol{\varphi})]^\top \bar{\boldsymbol{y}}_t = P_i^\lambda + P_i^\varphi$, where $P_i^\lambda = \bar{\boldsymbol{y}}_t^\top [\mathscr{D}_i \boldsymbol{S}(\boldsymbol{\varphi})] \boldsymbol{\Lambda}^{-2} (\mathrm{d}\boldsymbol{\Lambda}) [\boldsymbol{S}(\boldsymbol{\varphi})]^\top \bar{\boldsymbol{y}}_t$ and

$$P_i^\varphi = -\sum_{\ell=2}^k \Big(\bar{\boldsymbol{y}}_t^\top [\mathscr{D}_\ell \mathscr{D}_i \boldsymbol{S}(\boldsymbol{\varphi})] \boldsymbol{\Lambda}^{-1} [\boldsymbol{S}(\boldsymbol{\varphi})]^\top \bar{\boldsymbol{y}}_t + \bar{\boldsymbol{y}}_t^\top [\mathscr{D}_i \boldsymbol{S}(\boldsymbol{\varphi})] \boldsymbol{\Lambda}^{-1} [\mathscr{D}_\ell \boldsymbol{S}(\boldsymbol{\varphi})]^\top \bar{\boldsymbol{y}}_t \Big) \mathrm{d}\varphi_\ell.$$

We then obtain that $\mathrm{D}_{\boldsymbol{\varphi}} \boldsymbol{s}_t^\varphi(\boldsymbol{f})$ is a $(k-1)\times(k-1)$ matrix whose (i,j)th entry equals $-\bar{\boldsymbol{y}}_t^\top [\mathscr{D}_{j+1} \mathscr{D}_{i+1} \boldsymbol{S}(\boldsymbol{\varphi})] \boldsymbol{\Lambda}^{-1} [\boldsymbol{S}(\boldsymbol{\varphi})]^\top \bar{\boldsymbol{y}}_t - \bar{\boldsymbol{y}}_t^\top [\mathscr{D}_{i+1} \boldsymbol{S}(\boldsymbol{\varphi})] \boldsymbol{\Lambda}^{-1} [\mathscr{D}_{j+1} \boldsymbol{S}(\boldsymbol{\varphi})]^\top \bar{\boldsymbol{y}}_t$ for $1 \leqslant i,j \leqslant k-1$, and $\mathrm{D}_{\boldsymbol{\lambda}} \boldsymbol{s}_t^\varphi(\boldsymbol{f}) = \begin{pmatrix} [\bar{\boldsymbol{y}}_t^\top [\mathscr{D}_2 \boldsymbol{S}(\boldsymbol{\varphi})] \boldsymbol{\Lambda}^{-2}] \odot [\bar{\boldsymbol{y}}_t^\top \boldsymbol{S}(\boldsymbol{\varphi})] \\ \vdots \\ [\bar{\boldsymbol{y}}_t^\top [\mathscr{D}_k \boldsymbol{S}(\boldsymbol{\varphi})] \boldsymbol{\Lambda}^{-2}] \odot [\bar{\boldsymbol{y}}_t^\top \boldsymbol{S}(\boldsymbol{\varphi})] \end{pmatrix}$. Note that the spectral norm $\|\cdot\| \leqslant \|\cdot\|_F$, where $\|\cdot\|_F$ denotes the Frobenius norm. Moreover, for any compatible vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, we have $\|\boldsymbol{a} \odot \boldsymbol{b}\| = \|\mathrm{diag}(\boldsymbol{a})\boldsymbol{b}\| \leqslant \|\boldsymbol{a}\|\,\|\boldsymbol{b}\|$. Putting these results together, we obtain

$$\begin{aligned} Q_{1,t}(\alpha) &\leqslant \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sup_{\boldsymbol{f}\in\mathbb{F}} \left\{ \left\| \boldsymbol{B}_\lambda - \boldsymbol{A}_\lambda \right\| + \alpha \left\| \boldsymbol{A}_\varphi \right\| \left\| \mathrm{D}_{\boldsymbol{\lambda}} \boldsymbol{s}_t^\varphi(\boldsymbol{f}) \right\| \right\} \\ &\leqslant \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sup_{\boldsymbol{f}\in\mathbb{F}} \left\{ \left\| \boldsymbol{B}_\lambda - \boldsymbol{A}_\lambda \right\| + \alpha \left\| \boldsymbol{A}_\varphi \right\| \left(\sum_{i=2}^k \left\| [\bar{\boldsymbol{y}}_t^\top [\mathscr{D}_i \boldsymbol{S}(\boldsymbol{\varphi})] \boldsymbol{\Lambda}^{-2}] \odot [\bar{\boldsymbol{y}}_t^\top \boldsymbol{S}(\boldsymbol{\varphi})] \right\|_F^2 \right)^{1/2} \right\} \end{aligned}$$

$$\leqslant \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sup_{\boldsymbol{f}\in\mathbb{F}} \Big\{ \big\|\boldsymbol{B}_\lambda - \boldsymbol{A}_\lambda\big\| + \alpha\,\|\boldsymbol{A}_\varphi\|\,\sqrt{k-1}\,\|\bar{\boldsymbol{y}}_t\|\,\|\bar{\boldsymbol{y}}_t\|\,\big\|\boldsymbol{\Lambda}^{-2}\big\| \Big\}$$

$$\leqslant \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \Big\{ \big\|\boldsymbol{B}_\lambda - \boldsymbol{A}_\lambda\big\| + \alpha\,\|\boldsymbol{A}_\varphi\|\,\lambda_L^{-2}\sqrt{k-1}\,\|\boldsymbol{y}_t\|^2 \Big\},$$

and similarly, $Q_{2,t}(\alpha) \leqslant \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \big\{ \|\boldsymbol{B}_\varphi\| + 2\,\|\boldsymbol{A}_\varphi\|\,\lambda_L^{-1}(k-1)\,\|\boldsymbol{y}_t\|^2 + 2\,\alpha^{-1}\,\|\boldsymbol{A}_\lambda\|\,\sqrt{k-1}\,\|\boldsymbol{y}_t\|^2 \big\}$. Under Assumption 1(ii), it follows immediately that $\mathbb{E}\big[\log\rho(\boldsymbol{\mu}_t)\big] < 0$.

Then, by Theorem 2.8 of Straumann and Mikosch (2006), there exists a unique SE sequence $\big\{\boldsymbol{f}_t(\cdot),\, t\in\mathbb{Z}\big\}$ such that $\big\|\hat{\boldsymbol{f}}_t - \boldsymbol{f}_t\big\|_{\boldsymbol{\Theta},\alpha} \xrightarrow{e.a.s.} 0$ as $t\to\infty$. The first part of Proposition 3 then follows from the norm equivalence established in Lemma B.1. Moreover, the same theorem also implies that $\boldsymbol{f}_t$ is $\mathscr{F}_{t-1}$-measurable and admits the almost sure representation $\boldsymbol{f}_t = \lim_{r\to\infty}\boldsymbol{\mu}_{t-1}\circ\cdots\circ\boldsymbol{\mu}_{t-r}(\hat{\boldsymbol{f}}_1)$ for all $t\in\mathbb{Z}$, where the limit is independent of $\hat{\boldsymbol{f}}_1$ and the convergence holds in the norm $\|\cdot\|_{\boldsymbol{\Theta},\alpha}$. Moreover, for each $r\geqslant 1$, $\boldsymbol{\mu}_{t-1}\circ\cdots\circ\boldsymbol{\mu}_{t-r}(\hat{\boldsymbol{f}}_1)$ is continuous on $\boldsymbol{\Theta}$. Since $\boldsymbol{\Theta}$ is compact, it follows from the classical result that the uniform limit of continuous functions is continuous (Rudin, 1976, Theorem 7.12) that $\boldsymbol{f}_t$ is uniformly continuous on $\boldsymbol{\Theta}$; see also Lin and Lucas (2025, Proof of Proposition 1). □

Lemma A.1. *Recall* $\widehat{\mathcal{L}}_T(\boldsymbol{\theta}) = T^{-1}\sum_{t=1}^T \ell_t\big(\hat{\boldsymbol{f}}_t(\boldsymbol{\theta}),\boldsymbol{\nu}\big)$ *from* (2.8)*. Define the stationary counterpart* $\mathcal{L}_T(\boldsymbol{\theta}) = T^{-1}\sum_{t=1}^T \ell_t\big(\boldsymbol{f}_t(\boldsymbol{\theta}),\boldsymbol{\nu}\big)$ *and* $\mathcal{L}(\boldsymbol{\theta}) = \mathbb{E}\big(\ell_t(\boldsymbol{f}_t(\boldsymbol{\theta}),\boldsymbol{\nu})\big)$*. For the Gaussian specification in* (2.4)*, suppose that Assumption 1 and* $\sup_{t\in\mathbb{Z}}\mathbb{E}\,\|\boldsymbol{y}_t\|^2 < \infty$ *hold, we then have*

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\big|\widehat{\mathcal{L}}_T(\boldsymbol{\theta}) - \mathcal{L}(\boldsymbol{\theta})\big| \xrightarrow{a.s.} 0, \qquad T\to\infty, \tag{A.9}$$

where $\mathcal{L}(\cdot)$ *is continuous on* $\boldsymbol{\Theta}$*.*

Proof of Lemma A.1. The key steps of the proof are analogous to those in Lin et al. (2025, Lemma 1). We shall prove that ***(i)*** $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\big|\widehat{\mathcal{L}}_T(\boldsymbol{\theta}) - \mathcal{L}_T(\boldsymbol{\theta})\big| \xrightarrow{a.s.} 0$ and ***(ii)*** $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\big|\mathcal{L}_T(\boldsymbol{\theta}) - \mathcal{L}(\boldsymbol{\theta})\big| \xrightarrow{a.s.} 0$, as $T\to\infty$. Then (A.9) follows immediately from the triangle inequality. Note that, in the Gaussian model specification (2.4), $\ell_t(\boldsymbol{f},\boldsymbol{\nu})$ does not depend on $\boldsymbol{\nu}$ and can therefore be denoted simply by $\ell_t(\boldsymbol{f})$ below.

For the first term ***(i)***, by Proposition 3, there exists $\gamma > 1$ such that $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\hat{\boldsymbol{f}}_t(\boldsymbol{\theta}) - \boldsymbol{f}_t(\boldsymbol{\theta})\| \leqslant C\gamma^{-t}$ a.s. $\forall t \in \mathbb{Z}^+$. Applying a mean value theorem of $\ell_t\big(\hat{\boldsymbol{f}}_t(\boldsymbol{\theta})\big)$ around $\ell_t(\boldsymbol{f}_t(\boldsymbol{\theta}))$, we have

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \big|\widehat{\mathcal{L}}_T(\boldsymbol{\theta}) - \mathcal{L}_T(\boldsymbol{\theta})\big| \leqslant CT^{-1}\sum_{t=1}^{T}\left\{\gamma^{-t}\sup_{\boldsymbol{f}\in\mathbb{F}}\left\|\frac{\ell_t(\boldsymbol{f})}{\partial \boldsymbol{f}}\right\|\right\}, \tag{A.10}$$

where $\ell_t(\boldsymbol{f})/\partial\boldsymbol{f}$ is provided in Eq. (2.7). Given the SE property of $\boldsymbol{y}_t$, the separability of $\mathbb{F}$, and the continuity of the mapping $\boldsymbol{f} \mapsto \ell_t(\boldsymbol{f})$, it follows that $\big\{\sup_{\boldsymbol{f}\in\mathbb{F}} \|\partial\ell_t(\boldsymbol{f})/\partial\boldsymbol{f}\|, t \in \mathbb{Z}\big\}$ is SE. By the construction in (2.7), and following steps analogous to those in the proof of Proposition 3, it is straightforward to show that $\mathbb{E}\big(\log^+ \sup_{\boldsymbol{f}\in\mathbb{F}} \|\partial\ell_t(\boldsymbol{f})/\partial\boldsymbol{f}\|\big) < \infty$ provided that $\lambda_L > 0$ and $\mathbb{E}\big(\log^+\|\boldsymbol{y}_t\|\big) < \infty$, as already required by Assumption 1. The conclusion $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\big|\widehat{\mathcal{L}}_T(\boldsymbol{\theta}) - \mathcal{L}_T(\boldsymbol{\theta})\big| \xrightarrow{a.s.} 0$ then follows from Berkes et al. (2003, Lemma 2.2).

The second term ***(ii)*** and the continuity of $\boldsymbol{\theta} \mapsto \mathcal{L}(\boldsymbol{\theta})$ follow immediately from the uniform law of large numbers (ULLN) in White (1996, Theorem A.2.2), applied to the SE sequence $\big\{\ell_t(\boldsymbol{f}_t(\boldsymbol{\theta})), t \in \mathbb{Z}\big\}$ (by Proposition 3), provided that $\mathbb{E}\big(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\ell_t(\boldsymbol{f}_t(\boldsymbol{\theta}))|\big) < \infty$. By the definition (2.4) and Assumption 1(i), we have

$$\mathbb{E}\Big(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}|\ell_t(\boldsymbol{f}_t(\boldsymbol{\theta}))|\Big) \leqslant C + \frac{k}{2}\mathbb{E}\big(\log^+\|\boldsymbol{\lambda}_t\|_{\boldsymbol{\Theta}}\big) + \frac{1}{2}\lambda_L^{-1}\,\mathbb{E}\,\|\boldsymbol{y}_t\|^2. \tag{A.11}$$

Given the assumption that $\mathbb{E}\,\|\boldsymbol{y}_t\|^2 < \infty$, it suffices to show that $\mathbb{E}\big(\|\boldsymbol{\lambda}_t\|_{\boldsymbol{\Theta}}^p\big) < \infty$ for some $p \in (0,1]$ in order to establish (A.11). Let $K_p := \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\|\boldsymbol{B}_\lambda - \boldsymbol{A}_\lambda\|^p$. By one of the contraction conditions in Assumption 1(ii), it follows that $K_p < 1$ for any $p \in (0,1]$. By iterating backward $m \in \mathbb{Z}^+$ steps and applying the c_r-inequality for

$p \in (0,1]$, we arrive at

$$\mathbb{E}\big(\,\|\boldsymbol{\lambda}_t\|_{\boldsymbol{\Theta}}^p\,\big) \leqslant K_p^m\,\mathbb{E}\big(\,\|\boldsymbol{\lambda}_{t-m}\|_{\boldsymbol{\Theta}}^p\,\big) + \sum_{j=0}^{m-1} K_p^j \Big(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\boldsymbol{\omega}_\lambda\|^p + \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\boldsymbol{A}_\lambda\|^p\ \mathbb{E}\,\|\boldsymbol{y}_1\|^{2p}\Big). \quad \text{(A.12)}$$

For any $p \in (0,1]$, because $\{\boldsymbol{\lambda}_t,\, t\in\mathbb{Z}\}$ is SE, because $\mathbb{E}\,\|\boldsymbol{y}_1\|^{2p} < \infty$ follows from the condition $\sup_{t\in\mathbb{Z}} \mathbb{E}\,\|\boldsymbol{y}_t\|^2 < \infty$, and because $\boldsymbol{\Theta}$ is compact, we obtain $\mathbb{E}\big(\,\|\boldsymbol{\lambda}_t\|_{\boldsymbol{\Theta}}^p\,\big) \leqslant (1-K_p)^{-1}\big(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\boldsymbol{\omega}_\lambda\|^p + \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\boldsymbol{A}_\lambda\|^p\ \mathbb{E}\,\|\boldsymbol{y}_1\|^{2p}\big) < \infty$. This completes the proof. □

To proceed, we define

$$\begin{aligned} E_t :=& \big\{\lambda_{i,t}(\boldsymbol{\theta}_0) \neq \lambda_{j,t}(\boldsymbol{\theta}_0),\ \forall i \neq j,\ 1 \leqslant i,j \leqslant k\big\} \\ &\bigcap \Big\{\sin\big(\varphi_{k,t}(\boldsymbol{\theta}_0)\big) \neq 0,\ \cos\big(\varphi_{i,t}(\boldsymbol{\theta}_0)\big) \neq 0,\ 2 \leqslant i \leqslant k\Big\}, \quad t \in \mathbb{Z}. \end{aligned} \quad \text{(A.13)}$$

The event E_t rules out cases in which the eigenvalues are tied or the rotation angles take values at which the column structure of $\boldsymbol{S}(\boldsymbol{\varphi}_t)$, as discussed in Remark 1, fails to hold. Lemma B.3(iii) shows that E_t occurs a.s. for every $t \in \mathbb{Z}$ under Assumption 2(i).

Proof of Proposition 4. In the proof below, whenever a quantity depends on the static parameters $\boldsymbol{\theta}$, we make this dependence explicit. For instance, $\boldsymbol{V}_t = \bar{\boldsymbol{V}}\boldsymbol{S}(\boldsymbol{\varphi}_t)$ depends on $\boldsymbol{\theta}$ through $\boldsymbol{\varphi}_t = \boldsymbol{\varphi}_t(\boldsymbol{\theta})$. Accordingly, we write $\boldsymbol{V}_t(\boldsymbol{\theta})$ below whenever needed. Let $\boldsymbol{\Xi}_t(\boldsymbol{\theta},\boldsymbol{\theta}_0) := \boldsymbol{S}\big(\boldsymbol{\varphi}_t(\boldsymbol{\theta})\big)^\top \boldsymbol{S}\big(\boldsymbol{\varphi}_t(\boldsymbol{\theta}_0)\big)$. If $\boldsymbol{\Sigma}_t(\boldsymbol{\theta}) = \boldsymbol{\Sigma}_t(\boldsymbol{\theta}_0)$, then $\boldsymbol{\Lambda}_t(\boldsymbol{\theta}) = \boldsymbol{\Xi}_t(\boldsymbol{\theta},\boldsymbol{\theta}_0)\boldsymbol{\Lambda}_t(\boldsymbol{\theta}_0)\boldsymbol{\Xi}_t(\boldsymbol{\theta},\boldsymbol{\theta}_0)^\top$. Note that $\boldsymbol{\Xi}_t(\boldsymbol{\theta},\boldsymbol{\theta}_0)$ is an orthogonal matrix. Hence, the matrices $\boldsymbol{\Lambda}_t(\boldsymbol{\theta})$ and $\boldsymbol{\Lambda}_t(\boldsymbol{\theta}_0)$ are similar and share the same set of eigenvalues. Since they are diagonal, it follows that $\boldsymbol{\Lambda}_t(\boldsymbol{\theta})$ must be a permutation of $\boldsymbol{\Lambda}_t(\boldsymbol{\theta}_0)$. Note that the entries $\lambda_{i,t}(\boldsymbol{\theta}_0)$ are strictly positive. Moreover, they are distinct for $i = 1,\ldots,k$ when E_t occurs. By Lemma B.3(iii), it follows that $\boldsymbol{\Xi}_t(\boldsymbol{\theta},\boldsymbol{\theta}_0)$ must be a signed permutation matrix a.s. That is, it can be written as $\boldsymbol{\Xi}_t(\boldsymbol{\theta},\boldsymbol{\theta}_0) = \boldsymbol{\Pi}_t\boldsymbol{D}_t$, where $\boldsymbol{D}_t = \mathrm{diag}(\pm1,\pm1,\ldots,\pm1)$, and $\boldsymbol{\Pi}_t$ is a permutation matrix (i.e., a square matrix with exactly one entry equal to 1 in each row and each column, and all other entries equal to 0). This implies that $\boldsymbol{S}\big(\boldsymbol{\varphi}_t(\boldsymbol{\theta})\big) = \boldsymbol{S}\big(\boldsymbol{\varphi}_t(\boldsymbol{\theta}_0)\big)\boldsymbol{D}_t\boldsymbol{\Pi}_t^\top$. By Lemma B.4, $\boldsymbol{\Pi}_t^\top = \boldsymbol{\Pi}_{1k}^{\delta_t}$ for all t, where $\boldsymbol{\Pi}_{1k} = \big(\boldsymbol{e}_k, \boldsymbol{e}_2, \boldsymbol{e}_3, \ldots, \boldsymbol{e}_{k-1}, \boldsymbol{e}_1\big)$, $\boldsymbol{e}_i$ denotes the ith canonical basis vector in $\mathbb{R}^k$, and $\delta_t \in \{0,1\}$. That is, $\boldsymbol{\Pi}_t = \boldsymbol{\Pi}_t^\top = \boldsymbol{\Pi}_{1k}$ if $\delta_t = 1$, and $\boldsymbol{\Pi}_t = \boldsymbol{I}_k$ if $\delta_t = 0$. This implies that $\boldsymbol{\Lambda}_t(\boldsymbol{\theta}) = \boldsymbol{\Pi}_{1k}^{\delta_t}\,\boldsymbol{\Lambda}_t(\boldsymbol{\theta}_0)\,\boldsymbol{\Pi}_{1k}^{\delta_t}$, and hence $\boldsymbol{\lambda}_t(\boldsymbol{\theta}) = \boldsymbol{\Pi}_{1k}^{\delta_t}\,\boldsymbol{\lambda}_t(\boldsymbol{\theta}_0)$. Similarly, $\boldsymbol{\lambda}_{t+1}(\boldsymbol{\theta}) = \boldsymbol{\Pi}_{1k}^{\delta_{t+1}}\,\boldsymbol{\lambda}_{t+1}(\boldsymbol{\theta}_0)$ for all $t \in \mathbb{Z}$. We now proceed to establish the results (i)–(iii) separately.

Proof of Proposition 4(i): Combining the results above and writing $\widetilde{\boldsymbol{B}}_\lambda = \boldsymbol{B}_\lambda - \boldsymbol{A}_\lambda$, by the construction of the Gaussian SRE defined in (2.3) and (2.7), we have

$$\begin{aligned}
\boldsymbol{\Pi}_{1k}^{\delta_{t+1}}\,\boldsymbol{\lambda}_{t+1}(\boldsymbol{\theta}_0) &= \boldsymbol{\lambda}_{t+1}(\boldsymbol{\theta}) \\
&= (\boldsymbol{I}_k - \boldsymbol{B}_\lambda)\boldsymbol{\omega}_\lambda + \widetilde{\boldsymbol{B}}_\lambda\boldsymbol{\lambda}_t(\boldsymbol{\theta}) + \boldsymbol{A}_\lambda\Big[\boldsymbol{S}\big(\boldsymbol{\varphi}_t(\boldsymbol{\theta})\big)^\top\bar{\boldsymbol{V}}^\top\boldsymbol{y}_t\Big]^{\odot 2} \\
&= (\boldsymbol{I}_k - \boldsymbol{B}_\lambda)\boldsymbol{\omega}_\lambda + \widetilde{\boldsymbol{B}}_\lambda\,\boldsymbol{\Pi}_{1k}^{\delta_t}\,\boldsymbol{\lambda}_t(\boldsymbol{\theta}_0) + \boldsymbol{A}_\lambda\Big[\boldsymbol{\Pi}_{1k}^{\delta_t}\boldsymbol{D}_t\boldsymbol{S}\big(\boldsymbol{\varphi}_t(\boldsymbol{\theta}_0)\big)^\top\bar{\boldsymbol{V}}^\top\boldsymbol{y}_t\Big]^{\odot 2} \\
&= (\boldsymbol{I}_k - \boldsymbol{B}_\lambda)\boldsymbol{\omega}_\lambda + \big(\widetilde{\boldsymbol{B}}_\lambda\,\boldsymbol{\Pi}_{1k}^{\delta_t}\big)\,\boldsymbol{\lambda}_t(\boldsymbol{\theta}_0) + \big(\boldsymbol{A}_\lambda\boldsymbol{\Pi}_{1k}^{\delta_t}\big)\big(\boldsymbol{V}_t(\boldsymbol{\theta}_0)^\top\boldsymbol{y}_t\big)^{\odot 2}. \qquad \text{(A.14)}
\end{aligned}$$

Since the true dynamics are given by $\boldsymbol{\lambda}_{t+1}(\boldsymbol{\theta}_0) = (\boldsymbol{I}_k - \boldsymbol{B}_{\lambda 0})\boldsymbol{\omega}_{\lambda 0} + \widetilde{\boldsymbol{B}}_{\lambda 0}\boldsymbol{\lambda}_t(\boldsymbol{\theta}_0) + \boldsymbol{A}_{\lambda 0}\big(\boldsymbol{V}_t(\boldsymbol{\theta}_0)^\top\boldsymbol{y}_t\big)^{\odot 2}$, it follows that

$$\begin{aligned}
\boldsymbol{\Pi}_{1k}^{\delta_{t+1}}\,\boldsymbol{\lambda}_{t+1}(\boldsymbol{\theta}_0) &= \boldsymbol{\Pi}_{1k}^{\delta_{t+1}}\,(\boldsymbol{I}_k - \boldsymbol{B}_{\lambda 0})\boldsymbol{\omega}_{\lambda 0} \\
&\quad + \big(\boldsymbol{\Pi}_{1k}^{\delta_{t+1}}\,\widetilde{\boldsymbol{B}}_{\lambda 0}\big)\boldsymbol{\lambda}_t(\boldsymbol{\theta}_0) + \big(\boldsymbol{\Pi}_{1k}^{\delta_{t+1}}\,\boldsymbol{A}_{\lambda 0}\big)\big(\boldsymbol{V}_t(\boldsymbol{\theta}_0)^\top\boldsymbol{y}_t\big)^{\odot 2}. \qquad \text{(A.15)}
\end{aligned}$$

Note that for any $t \in \mathbb{Z}$, Lemma B.3(iii) implies that $\mathbb{P}(E_t \cap E_{t+1}) = 1 - \mathbb{P}\big(E_t^\complement \cup E_{t+1}^\complement\big) = 1$. Fix $t \in \mathbb{Z}$ and set $(\delta, \delta') = (\delta_t, \delta_{t+1}) \in \{0,1\} \times \{0,1\}$. Then the pair (δ_t, δ_{t+1}) is fixed as (δ, δ') at this time index. By (A.14)–(A.15), we obtain an equality of the form (B.5) in Lemma B.3. Hence, by Assumption 2(i) and Lemma B.3(i), it follows a.s. that

$$(\boldsymbol{I}_k - \boldsymbol{B}_\lambda)\boldsymbol{\omega}_\lambda = \boldsymbol{\Pi}_{1k}^{\delta'}\,(\boldsymbol{I}_k - \boldsymbol{B}_{\lambda 0})\boldsymbol{\omega}_{\lambda 0}, \quad \widetilde{\boldsymbol{B}}_\lambda = \boldsymbol{\Pi}_{1k}^{\delta'}\,\widetilde{\boldsymbol{B}}_{\lambda 0}\,\boldsymbol{\Pi}_{1k}^{\delta}, \quad \boldsymbol{A}_\lambda = \boldsymbol{\Pi}_{1k}^{\delta'}\,\boldsymbol{A}_{\lambda 0}\,\boldsymbol{\Pi}_{1k}^{\delta}. \qquad \text{(A.16)}$$

By Assumption 2(ii), both $\widetilde{\boldsymbol{B}}_\lambda$ and $\widetilde{\boldsymbol{B}}_{\lambda 0}$ are diagonal, which forces the permutation to preserve diagonality and hence implies $\delta = \delta' \in \{0, 1\}$, given that the first and/or the last diagonal element of both $\widetilde{\boldsymbol{B}}_\lambda$ and $\widetilde{\boldsymbol{B}}_{\lambda 0}$ are non-zero. Since $\widetilde{\boldsymbol{B}}_\lambda = \boldsymbol{\Pi}_{1k}^{\delta}\, \widetilde{\boldsymbol{B}}_{\lambda 0}\, \boldsymbol{\Pi}_{1k}^{\delta}$ and $\boldsymbol{A}_\lambda = \boldsymbol{\Pi}_{1k}^{\delta}\, \boldsymbol{A}_{\lambda 0}\, \boldsymbol{\Pi}_{1k}^{\delta}$, it follows by summation that $\boldsymbol{B}_\lambda = \boldsymbol{\Pi}_{1k}^{\delta}\, \boldsymbol{B}_{\lambda 0}\, \boldsymbol{\Pi}_{1k}^{\delta}$ and

$$
\begin{aligned}
&(\boldsymbol{I}_k - \boldsymbol{B}_\lambda)\boldsymbol{\omega}_\lambda = \boldsymbol{\Pi}_{1k}^{\delta}\, (\boldsymbol{I}_k - \boldsymbol{B}_{\lambda 0})\boldsymbol{\omega}_{\lambda 0} \\
&\quad\iff \quad (\boldsymbol{I}_k - \boldsymbol{\Pi}_{1k}^{\delta}\, \boldsymbol{B}_{\lambda 0}\, \boldsymbol{\Pi}_{1k}^{\delta})\boldsymbol{\omega}_\lambda = \boldsymbol{\Pi}_{1k}^{\delta}\, (\boldsymbol{I}_k - \boldsymbol{B}_{\lambda 0})\boldsymbol{\omega}_{\lambda 0} \\
&\quad\iff \quad \boldsymbol{\Pi}_{1k}^{\delta}(\boldsymbol{I}_k - \boldsymbol{\Pi}_{1k}^{\delta}\, \boldsymbol{B}_{\lambda 0}\, \boldsymbol{\Pi}_{1k}^{\delta})\boldsymbol{\omega}_\lambda = (\boldsymbol{I}_k - \boldsymbol{B}_{\lambda 0})\boldsymbol{\omega}_{\lambda 0} \\
&\quad\iff \quad (\boldsymbol{I}_k - \boldsymbol{B}_{\lambda 0})\boldsymbol{\Pi}_{1k}^{\delta}\, \boldsymbol{\omega}_\lambda = (\boldsymbol{I}_k - \boldsymbol{B}_{\lambda 0})\boldsymbol{\omega}_{\lambda 0}.
\end{aligned}
$$

Since $\|\boldsymbol{B}_{\lambda 0}\| < 1$ (Assumption 2(ii)), $\boldsymbol{I}_k - \boldsymbol{B}_{\lambda 0}$ is invertible. It then follows that $\boldsymbol{\omega}_\lambda = \boldsymbol{\Pi}_{1k}^{\delta}\, \boldsymbol{\omega}_{\lambda 0}$. Let $\omega_{\lambda 0, i}$ be the ith element of $\boldsymbol{\omega}_{\lambda 0}$ for $i = 1, \ldots, k$. Imposing the normalization $\omega_{\lambda,1} > \omega_{\lambda,k}$ and $\omega_{\lambda 0,1} > \omega_{\lambda 0,k}$ then rules out the case $\delta = 1$, and therefore yields $(\boldsymbol{\omega}_\lambda,\, \boldsymbol{B}_\lambda,\, \boldsymbol{A}_\lambda) = (\boldsymbol{\omega}_{\lambda 0},\, \boldsymbol{B}_{\lambda 0},\, \boldsymbol{A}_{\lambda 0})$ and $\boldsymbol{\Xi}_t(\boldsymbol{\theta}, \boldsymbol{\theta}_0) = \boldsymbol{D}_t$, which implies that $\boldsymbol{S}\big(\boldsymbol{\varphi}_t(\boldsymbol{\theta})\big) = \boldsymbol{S}\big(\boldsymbol{\varphi}_t(\boldsymbol{\theta}_0)\big)\boldsymbol{D}_t$ for every $t \in \mathbb{Z}$ a.s.

Proof of Proposition 4(ii): By the preceding arguments and Lemma B.6, we immediately obtain Proposition 4(ii), where $\boldsymbol{D}_t = \boldsymbol{D}(\boldsymbol{q}_t)$. Here, for an integer vector $\boldsymbol{q} = (q_2, \ldots, q_k)^\top \in \mathbb{Z}^{k-1}$, $\boldsymbol{D}(\boldsymbol{q}) := \prod_{i=2}^{k} \boldsymbol{D}_{1i}^{q_i}$ denotes a $k \times k$ diagonal sign matrix, where $\boldsymbol{D}_{1i}$ is an identity matrix with the $(1,1)$ and (i,i) diagonal entries replaced by -1.

Proof of Proposition 4(iii): By the representations in Proposition 4(ii) as derived above and the identities (I1)–(I2) given in the proof of Lemma B.5, applied repeatedly, we obtain, for $i = 2, \ldots, k$,

$$
\begin{aligned}
&\boldsymbol{y}_t^\top \bar{\boldsymbol{V}}[\mathscr{D}_i \boldsymbol{S}(\boldsymbol{\varphi}_t(\boldsymbol{\theta}))]\boldsymbol{\Lambda}_t(\boldsymbol{\theta})^{-1}\boldsymbol{V}_t(\boldsymbol{\theta})^\top \boldsymbol{y}_t \\
&\quad = \boldsymbol{y}_t^\top \bar{\boldsymbol{V}}[\mathscr{D}_i \boldsymbol{S}(\boldsymbol{\varphi}_t(\boldsymbol{\theta}))]\boldsymbol{\Lambda}_t(\boldsymbol{\theta})^{-1}\boldsymbol{S}(\boldsymbol{\varphi}_t(\boldsymbol{\theta}))^\top \bar{\boldsymbol{V}}^\top \boldsymbol{y}_t \\
&\quad = \zeta_i(\boldsymbol{q}_t)\boldsymbol{y}_t^\top \bar{\boldsymbol{V}}[\mathscr{D}_i \boldsymbol{S}(\boldsymbol{\varphi}_t(\boldsymbol{\theta}_0))]\, \boldsymbol{D}(\boldsymbol{q}_t)\boldsymbol{\Lambda}_t(\boldsymbol{\theta}_0)^{-1}\boldsymbol{D}(\boldsymbol{q}_t)\, \boldsymbol{S}\big(\boldsymbol{\varphi}_t(\boldsymbol{\theta}_0)\big)^\top \bar{\boldsymbol{V}}^\top \boldsymbol{y}_t
\end{aligned}
$$

$$= \zeta_i(\boldsymbol{q}_t)\, \boldsymbol{y}_t^\top \bar{\boldsymbol{V}}[\mathscr{D}_i \boldsymbol{S}(\boldsymbol{\varphi}_t(\boldsymbol{\theta}_0))]\, \boldsymbol{\Lambda}_t(\boldsymbol{\theta}_0)^{-1} \boldsymbol{V}_t(\boldsymbol{\theta}_0)^\top \boldsymbol{y}_t, \tag{A.17}$$

where we use the fact that $\boldsymbol{\lambda}_t(\boldsymbol{\theta}) = \boldsymbol{\lambda}_t(\boldsymbol{\theta}_0)$ a.s. by the proof of Part (i). Using (A.17) and the likelihood score in (2.7), we can write

$$\left.\frac{\partial \ell_t(\boldsymbol{f}_t, \boldsymbol{\nu})}{\partial \boldsymbol{\varphi}_t}\right|_{\boldsymbol{\varphi}_t = \boldsymbol{\zeta}(\boldsymbol{q}_t)\, \boldsymbol{\varphi}_t(\boldsymbol{\theta}_0) + \pi\, \boldsymbol{q}_t,\ \boldsymbol{\lambda}_t = \boldsymbol{\lambda}_t(\boldsymbol{\theta}_0)} = \boldsymbol{\zeta}(\boldsymbol{q}_t) \cdot \left.\frac{\partial \ell_t(\boldsymbol{f}_t, \boldsymbol{\nu})}{\partial \boldsymbol{\varphi}_t}\right|_{\boldsymbol{\varphi}_t = \boldsymbol{\varphi}_t(\boldsymbol{\theta}_0),\ \boldsymbol{\lambda}_t = \boldsymbol{\lambda}_t(\boldsymbol{\theta}_0)}. \tag{A.18}$$

Therefore, by the representations given in Proposition 4(ii) and (A.18), and by the construction of the Gaussian SRE defined in (2.3) and (2.7), we arrive at

$$\begin{aligned} \boldsymbol{\zeta}(\boldsymbol{q}_{t+1})\, \boldsymbol{\varphi}_{t+1}(\boldsymbol{\theta}_0) + \pi\, \boldsymbol{q}_{t+1} &= \boldsymbol{\varphi}_{t+1}(\boldsymbol{\theta}) = (\boldsymbol{I}_{k-1} - \boldsymbol{B}_\varphi)\boldsymbol{\omega}_\varphi \\ &\quad + \boldsymbol{B}_\varphi\big[\boldsymbol{\zeta}(\boldsymbol{q}_t)\, \boldsymbol{\varphi}_t(\boldsymbol{\theta}_0) + \pi\, \boldsymbol{q}_t\big] + \boldsymbol{A}_\varphi\, \boldsymbol{\zeta}(\boldsymbol{q}_t) \left.\frac{\partial \ell_t(\boldsymbol{f}_t, \boldsymbol{\nu})}{\partial \boldsymbol{\varphi}_t}\right|_{\boldsymbol{\varphi}_t = \boldsymbol{\varphi}_t(\boldsymbol{\theta}_0),\ \boldsymbol{\lambda}_t = \boldsymbol{\lambda}_t(\boldsymbol{\theta}_0)}. \end{aligned} \tag{A.19}$$

On the other hand, using the true dynamics, we have

$$\begin{aligned} \boldsymbol{\zeta}(\boldsymbol{q}_{t+1})\, \boldsymbol{\varphi}_{t+1}(\boldsymbol{\theta}_0) &= \boldsymbol{\zeta}(\boldsymbol{q}_{t+1})(\boldsymbol{I}_{k-1} - \boldsymbol{B}_{\varphi 0})\boldsymbol{\omega}_{\varphi 0} \\ &\quad + \boldsymbol{\zeta}(\boldsymbol{q}_{t+1})\boldsymbol{B}_{\varphi 0}\, \boldsymbol{\varphi}_t(\boldsymbol{\theta}_0) + \boldsymbol{\zeta}(\boldsymbol{q}_{t+1})\boldsymbol{A}_{\varphi 0} \left.\frac{\partial \ell_t(\boldsymbol{f}_t, \boldsymbol{\nu})}{\partial \boldsymbol{\varphi}_t}\right|_{\boldsymbol{\varphi}_t = \boldsymbol{\varphi}_t(\boldsymbol{\theta}_0),\ \boldsymbol{\lambda}_t = \boldsymbol{\lambda}_t(\boldsymbol{\theta}_0)}. \end{aligned} \tag{A.20}$$

Comparing (A.19) and (A.20), one has

$$\begin{aligned} &\boldsymbol{\kappa}(\boldsymbol{q}_t, \boldsymbol{q}_{t+1}) + \boldsymbol{G}(\boldsymbol{q}_t, \boldsymbol{q}_{t+1})\, \boldsymbol{\varphi}_t(\boldsymbol{\theta}_0) \\ &\qquad\qquad + \boldsymbol{H}(\boldsymbol{q}_t, \boldsymbol{q}_{t+1}) \left.\frac{\partial \ell_t(\boldsymbol{f}_t, \boldsymbol{\nu})}{\partial \boldsymbol{\varphi}_t}\right|_{\boldsymbol{\varphi}_t = \boldsymbol{\varphi}_t(\boldsymbol{\theta}_0),\ \boldsymbol{\lambda}_t = \boldsymbol{\lambda}_t(\boldsymbol{\theta}_0)} = \boldsymbol{0}, \end{aligned}$$

where $\boldsymbol{\kappa}(\boldsymbol{q}_t, \boldsymbol{q}_{t+1}) := (\boldsymbol{I}_{k-1} - \boldsymbol{B}_\varphi)\boldsymbol{\omega}_\varphi - \boldsymbol{\zeta}(\boldsymbol{q}_{t+1})(\boldsymbol{I}_{k-1} - \boldsymbol{B}_{\varphi 0})\boldsymbol{\omega}_{\varphi 0} + \pi\, \boldsymbol{B}_\varphi\, \boldsymbol{q}_t - \pi\, \boldsymbol{q}_{t+1}$, $\boldsymbol{G}(\boldsymbol{q}_t, \boldsymbol{q}_{t+1}) := \boldsymbol{B}_\varphi \boldsymbol{\zeta}(\boldsymbol{q}_t) - \boldsymbol{\zeta}(\boldsymbol{q}_{t+1})\boldsymbol{B}_{\varphi 0}$, and $\boldsymbol{H}(\boldsymbol{q}_t, \boldsymbol{q}_{t+1}) := \boldsymbol{A}_\varphi \boldsymbol{\zeta}(\boldsymbol{q}_t) - \boldsymbol{\zeta}(\boldsymbol{q}_{t+1})\boldsymbol{A}_{\varphi 0}$. For the argument below, take an arbitrary representation obtained in Part (ii) as fixed. For each fixed $t \in \mathbb{Z}$, Lemma B.3 is then applied to the resulting fixed pair $(\boldsymbol{q}_t, \boldsymbol{q}_{t+1})$. By Lemma B.3(ii), this implies that $\boldsymbol{G}(\boldsymbol{q}_t, \boldsymbol{q}_{t+1}) = \boldsymbol{H}(\boldsymbol{q}_t, \boldsymbol{q}_{t+1}) = \boldsymbol{0}$ and

$\boldsymbol{\kappa}(\boldsymbol{q}_t, \boldsymbol{q}_{t+1}) = \mathbf{0}$.

Given the diagonality of $\boldsymbol{B}_\varphi$ and $\boldsymbol{B}_{\varphi 0}$, we obtain $\boldsymbol{B}_\varphi = \boldsymbol{\zeta}(\boldsymbol{q}_{t+1})\boldsymbol{B}_{\varphi 0}\,\boldsymbol{\zeta}(\boldsymbol{q}_t) = \boldsymbol{B}_{\varphi 0}\,\bar{\boldsymbol{\zeta}}_t$ by $\boldsymbol{G}(\boldsymbol{q}_t, \boldsymbol{q}_{t+1}) = \mathbf{0}$, where $\bar{\boldsymbol{\zeta}}_t = \mathrm{diag}(\bar{\zeta}_{2,t}, \ldots, \bar{\zeta}_{k,t}) = \boldsymbol{\zeta}(\boldsymbol{q}_{t+1})\boldsymbol{\zeta}(\boldsymbol{q}_t)$ with ± 1s on the diagonal. From $\boldsymbol{\kappa}(\boldsymbol{q}_t, \boldsymbol{q}_{t+1}) = \mathbf{0}$, we also obtain

$$\boldsymbol{q}_{t+1} = \boldsymbol{B}_{\varphi 0}\,\bar{\boldsymbol{\zeta}}_t\,\boldsymbol{q}_t + \frac{1}{\pi}\big[(\boldsymbol{I}_{k-1} - \boldsymbol{B}_{\varphi 0}\,\bar{\boldsymbol{\zeta}}_t)\boldsymbol{\omega}_\varphi - (\boldsymbol{I}_{k-1} - \boldsymbol{B}_{\varphi 0})\boldsymbol{\zeta}(\boldsymbol{q}_{t+1})\boldsymbol{\omega}_{\varphi 0}\big]. \tag{A.21}$$

For the fixed representation under consideration, (A.21) holds with $\boldsymbol{q}_t, \boldsymbol{q}_{t+1} \in \mathbb{Z}^{k-1}$ for all $t \in \mathbb{Z}$.

Before continuing, note that, for an integer-valued scalar sequence $\{r_t,\, t \in \mathbb{Z}\}$ satisfying $r_{t+1} = br_t + c$ with $|b| < 1$ and $c \in \mathbb{R}$, one must have $r_t = c/(1-b)$ for all $t \in \mathbb{Z}$. Indeed, if $b = 0$, then $r_{t+1} = c$ for all t, and hence $r_t = c = c/(1-b)$ for all t. If $b \neq 0$, fix $t \in \mathbb{Z}$ in what follows, and let $r^* = c/(1-b)$. Iterating the recursion gives $r_{t+n} - r^* = b^n(r_t - r^*)$ for all $n \geqslant 0$. Since $|b| < 1$, $r_{t+n} \to r^*$ as $n \to \infty$. Since $r_{t+n} \to r^*$ (as $n \to \infty$) and each r_{t+n} is an integer, the sequence $\{r_{t+n},\, n \geqslant 0\}$ must be eventually constant. Indeed, for all sufficiently large n, the terms r_{t+n} lie within distance $1/2$ of the limit r^*; an integer sequence can contain at most one value in such a neighborhood. Hence $r_{t+n} = r^*$ for all sufficiently large n. Since $r_{s+1} - r^* = b(r_s - r^*)$ and $b \neq 0$, $r_{s+1} = r^*$ implies $r_s = r^*$. Repeating this step backward from $s = t + n - 1$ down to $s = t$ gives $r_t = r^*$. Since t is arbitrary, the claim follows.

Now write $\boldsymbol{q}_t = (q_{2,t}, q_{3,t}, \ldots, q_{k,t})^\top$ and $\boldsymbol{B}_{\varphi 0} = \mathrm{diag}(B_{\varphi 0,2}, B_{\varphi 0,3}, \ldots, B_{\varphi 0,k})$. Note that $\bar{\zeta}_{2,t} = 1$ for all $t \in \mathbb{Z}$ because $\zeta_2(\boldsymbol{q}_t) = 1$ for every $\boldsymbol{q}_t \in \mathbb{Z}^{k-1}$. By the assumption $|B_{\varphi 0,2}| < 1$, regardless of whether $B_{\varphi 0,2}$ is zero or not, the first coordinate of (A.21) yields

$$q_{2,t+1} = B_{\varphi 0,2}\, q_{2,t} + \frac{(1 - B_{\varphi 0,2})(\omega_{\varphi,2} - \omega_{\varphi 0,2})}{\pi} \quad \Longrightarrow \quad q_{2,t} = \frac{\omega_{\varphi,2} - \omega_{\varphi 0,2}}{\pi} \in (-1, 1),$$

under the restriction $\omega_{\varphi,i} \in (\omega_L, \omega_L + \pi]$ for $i = 2, \ldots, k$ and some $\omega_L \in \mathbb{R}$ as imposed in Assumption 2(ii). Since $q_{2,t} \in \mathbb{Z}$, this implies that $q_{2,t} = 0$ for all $t \in \mathbb{Z}$, and thus $\omega_{\varphi,2} = \omega_{\varphi 0,2}$. We proceed by induction. Suppose that, for some $i \in \{3, \ldots, k\}$, $q_{2,t} = \cdots = q_{i-1,t} = 0$ for all $t \in \mathbb{Z}$. Then, by the structure of $\boldsymbol{\zeta}(\boldsymbol{q}_t)$, we have $\zeta_i(\boldsymbol{q}_t) = 1$ for all $t \in \mathbb{Z}$, and hence $\bar{\zeta}_{i,t} = 1$. Similarly, the ith coordinate of (A.21) gives

$$q_{i,t+1} = B_{\varphi 0,i}\, q_{i,t} + \frac{(1 - B_{\varphi 0,i})(\omega_{\varphi,i} - \omega_{\varphi 0,i})}{\pi} \qquad \Longrightarrow \qquad q_{i,t} = \frac{\omega_{\varphi,i} - \omega_{\varphi 0,i}}{\pi} \in (-1, 1).$$

Using the same argument as above, we obtain $q_{i,t} = 0$ for all $t \in \mathbb{Z}$ and thus $\omega_{\varphi,i} = \omega_{\varphi 0,i}$, where $i \in \{3, \ldots, k\}$. As a result, we obtain $\boldsymbol{q}_t = \mathbf{0}$ for all $t \in \mathbb{Z}$ and thus $\boldsymbol{\omega}_\varphi = \boldsymbol{\omega}_{\varphi 0}$. At the same time, we obtain $\boldsymbol{\zeta}(\boldsymbol{q}_t) = \boldsymbol{\zeta}(\boldsymbol{q}_{t+1}) = \boldsymbol{I}_{k-1}$, and thus $\boldsymbol{A}_\varphi = \boldsymbol{A}_{\varphi 0}$, $\boldsymbol{B}_\varphi = \boldsymbol{B}_{\varphi 0}$. This completes the proof. □

Proof of Theorem 1. The proof follows from standard consistency arguments as in Blasques et al. (2022, Theorem 4.6), which rely on Lemma A.1, Proposition 4, and White (1996, Theorem 3.4). □

Proof of Theorem 2. We prove the theorem by applying Theorem 3(b) of Andrews (1999). In view of Theorem 1 in the same reference, it suffices to verify Assumptions 1, 2*, 3*, 5*, and 6 therein. Recall that in the Gaussian filter dynamics, that is, under (2.4), $\ell_t(\boldsymbol{f}, \boldsymbol{\nu})$ does not depend on $\boldsymbol{\nu}$ and may therefore be written simply as $\ell_t(\boldsymbol{f})$ without confusion. Also recall the notation $\widehat{\mathcal{L}}_T(\boldsymbol{\theta}) = T^{-1} \sum_{t=1}^{T} \ell_t\big(\hat{\boldsymbol{f}}_t(\boldsymbol{\theta})\big)$ and $\mathcal{L}_T(\boldsymbol{\theta}) = T^{-1} \sum_{t=1}^{T} \ell_t(\boldsymbol{f}_t(\boldsymbol{\theta}))$ as given in Lemma A.1. Note that Assumption 1 of Andrews (1999) follows immediately from Theorem 1.

Verification of Assumption 2*. We first establish the condition in Eq. (9.1) of Andrews (2001), which in the current paper reduces to showing that

$$\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}:\, \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \leqslant \gamma_T} T \Big| \big(\widehat{\mathcal{L}}_T(\boldsymbol{\theta}) - \widehat{\mathcal{L}}_T(\boldsymbol{\theta}_0)\big) - \big(\mathcal{L}_T(\boldsymbol{\theta}) - \mathcal{L}_T(\boldsymbol{\theta}_0)\big) \Big| = o_{\mathbb{P}}(1) \tag{A.22}$$

for all $\gamma_T \to 0$. Note that the left-hand side of (A.22) is bounded by

$$\sum_{t=1}^{T} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}:\, \|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\leqslant\gamma_T} \left| \Big(\ell_t\big(\hat{\boldsymbol{f}}_t(\boldsymbol{\theta})\big) - \ell_t(\boldsymbol{f}_t(\boldsymbol{\theta}))\Big) - \Big(\ell_t\big(\hat{\boldsymbol{f}}_t(\boldsymbol{\theta}_0)\big) - \ell_t(\boldsymbol{f}_t(\boldsymbol{\theta}_0))\Big) \right|. \quad \text{(A.23)}$$

By the arguments given below (A.10), we have $\sum_{t=1}^{\infty} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left|\ell_t\big(\hat{\boldsymbol{f}}_t(\boldsymbol{\theta})\big) - \ell_t(\boldsymbol{f}_t(\boldsymbol{\theta}))\right| < \infty$. Moreover, $\ell_t\big(\hat{\boldsymbol{f}}_t(\boldsymbol{\theta})\big) - \ell_t(\boldsymbol{f}_t(\boldsymbol{\theta}))$ is continuous in $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ by Proposition 3. Hence, by (A.23) and by the same argument as in Eq. (9.26) of Andrews (2001), (A.22) is obtained immediately by further noting that $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}:\, \|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\leqslant\gamma_T} \left|\Big(\ell_t\big(\hat{\boldsymbol{f}}_t(\boldsymbol{\theta})\big) - \ell_t(\boldsymbol{f}_t(\boldsymbol{\theta}))\Big) - \Big(\ell_t\big(\hat{\boldsymbol{f}}_t(\boldsymbol{\theta}_0)\big) - \ell_t(\boldsymbol{f}_t(\boldsymbol{\theta}_0))\Big)\right| \leqslant 2 \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}:\, \|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\leqslant\gamma_T} \left|\ell_t\big(\hat{\boldsymbol{f}}_t(\boldsymbol{\theta})\big) - \ell_t(\boldsymbol{f}_t(\boldsymbol{\theta}))\right|$.

Following arguments similar to those in Andrews (2001, Appendix A), in particular Eqs. (9.2)–(9.3), one can obtain the quadratic approximation of $\widehat{\mathcal{L}}_T(\cdot)$ given in Eq. (3.2) of Andrews (1999) directly in terms of $\partial\mathcal{L}_T(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}$ and $\partial^2\mathcal{L}_T(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top$. That is, if Assumption 2^{2^*} in Andrews (1999) holds for $\mathcal{L}_T(\cdot)$, then Assumption 2^* holds for $\mathcal{L}_T(\cdot)$; in turn, by (A.22), Assumption 2^* also holds for $\widehat{\mathcal{L}}_T(\cdot)$, with the quadratic approximation in Eq. (3.2) of Andrews (1999) defined in terms of $\partial\mathcal{L}_T(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}$ and $\partial^2\mathcal{L}_T(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top$ for the terms involving the operators D and D^2, respectively.

Now we verify Assumption 2^{2^*} for $\mathcal{L}_T(\cdot)$. Let $\boldsymbol{\Theta}^+$ in Andrews (1999) be given by $\boldsymbol{\Theta}_0^{\mathbb{C},\epsilon} = \boldsymbol{\Theta} \cap \mathbb{C}(\boldsymbol{\theta}_0, \epsilon)$, where $\epsilon > 0$ is as specified in Assumption 3(ii). Then Assumption 2^{2^*}(a) is trivially satisfied given Assumption 3(i). Assumption 2^{2^*}(b) follows from Assumption 3(ii). To verify Assumption 2^{2^*}(c), it suffices to show that

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_0^\epsilon} \left\| \left(-\frac{\partial^2\mathcal{L}_T(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} \right) - \boldsymbol{\mathcal{J}}(\boldsymbol{\theta}) \right\| = o_{\mathbb{P}}(1), \quad \text{(A.24)}$$

where $\boldsymbol{\Theta}_0^\epsilon \subset \boldsymbol{\Theta}_0^{\mathbb{C},\epsilon}$ (see Assumption 3(iii) for the definition), and $\boldsymbol{\mathcal{J}}(\boldsymbol{\theta}) = -\mathbb{E}\Big(\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\ell_t(\boldsymbol{f}_t(\boldsymbol{\theta}))\Big)$. Note that $\partial^2\ell_t(\boldsymbol{f}_t(\boldsymbol{\theta}))/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top$ is a measurable function of $\big(\boldsymbol{y}_t, \boldsymbol{f}_t(\boldsymbol{\theta}), \partial\boldsymbol{f}_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}, \partial^2\boldsymbol{f}_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top\big)$. By Assumption 3(ii), it is continuous in $\boldsymbol{\theta} \in \boldsymbol{\Theta}_0^\epsilon$ and is SE as a process indexed by t. Therefore, (A.24) follows directly from

Assumption 3(iii) and the ULLN for SE processes in White (1996, Theorem A.2.2).

Verification of Assumption 3*. As discussed above, it suffices to verify Assumption 3* for $\widehat{\mathcal{L}}_T(\cdot)$ using the terms appearing in the quadratic approximation, namely $\partial \mathcal{L}_T(\boldsymbol{\theta}_0)/\partial \boldsymbol{\theta}$ and $\partial^2 \mathcal{L}_T(\boldsymbol{\theta}_0)/\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top$. We first establish the asymptotic distribution of $\sqrt{T}\, \partial \mathcal{L}_T(\boldsymbol{\theta}_0)/\partial \boldsymbol{\theta}$. Note that

$$\sqrt{T}\, \frac{\partial \mathcal{L}_T(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\partial \ell_t(\boldsymbol{f}_t(\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(\frac{\partial \boldsymbol{f}_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^\top} \right)^\top \frac{\partial \ell_t(\boldsymbol{f})}{\partial \boldsymbol{f}}\bigg|_{\boldsymbol{f}=\boldsymbol{f}_t(\boldsymbol{\theta}_0)}$$

where $\partial \boldsymbol{f}_t(\boldsymbol{\theta}_0)/\partial \boldsymbol{\theta}^\top$ is $\mathscr{F}_{t-1}$-measurable by Assumption 3(ii). Since $\boldsymbol{\varepsilon}_t \sim (\mathbf{0}, \boldsymbol{I}_k)$ are i.i.d. and $\{\boldsymbol{y}_t, t \in \mathbb{Z}\}$ is generated according to $\{\boldsymbol{y}_t^{\varepsilon}, t \in \mathbb{Z}\}$ in Proposition 2, it is not hard to compute that $\mathbb{E}\Big(\partial \ell_t(\boldsymbol{f})/\partial \boldsymbol{f}\, \big|_{\boldsymbol{f}=\boldsymbol{f}_t(\boldsymbol{\theta}_0)} \,\Big|\, \mathscr{F}_{t-1} \Big) = \mathbf{0}$ and thus $\sum_{t=1}^{T} \Big[\partial \ell_t(\boldsymbol{f}_t(\boldsymbol{\theta}))/\partial \boldsymbol{\theta}\, \big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \Big]$ is a partial sum of SE martingale difference (m.d.) random variables. We 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, as well as the discussion in Chapter 24.3, p. 385 of the same reference), provided that $\mathbb{E}\big(\big\| \partial \ell_t(\boldsymbol{f}_t(\boldsymbol{\theta}))/\partial \boldsymbol{\theta}\, \big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \big\|^2 \big) < \infty$ as required in Assumption 3(iii). This implies that, as $T \to \infty$,

$$\sqrt{T}\, \frac{\partial \mathcal{L}_T(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\mathcal{I}}_0), \qquad \boldsymbol{\mathcal{I}}_0 = \mathbb{E}\left(\frac{\partial \ell_t(\boldsymbol{f}_t(\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}} \frac{\partial \ell_t(\boldsymbol{f}_t(\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}^\top}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right). \tag{A.25}$$

Finally, note that $\boldsymbol{\mathcal{J}}(\boldsymbol{\theta}_0)$ is symmetric, nonrandom, and nonsingular by Assumption 3(iii). This completes the verification of Assumption 3*.

Verification of Assumptions 5* and 6. Since $\boldsymbol{\theta}_0 \in \mathrm{int}(\boldsymbol{\Theta})$, we may take the cone $\boldsymbol{\Lambda}$ in Assumption 5*(a) of Andrews (1999) to be $\boldsymbol{\Lambda} = \mathbb{R}^{\dim(\boldsymbol{\theta}_0)}$, where $\dim(\boldsymbol{\theta}_0)$ denotes the dimension of $\boldsymbol{\theta}_0$. Moreover, Assumption 5*(b) is satisfied by taking $b_T = \sqrt{T}$. Hence, Assumption 5* holds. Assumption 6 follows immediately because the cone $\boldsymbol{\Lambda}$ specified above is convex.

The proof is completed by a direct application of Theorem 3(b) of Andrews (1999). $\square$

B Proofs of auxiliary results

Lemma B.1 (Equivalence of $\|\cdot\|_{\boldsymbol{\Theta},\alpha}$ and $\|\cdot\|_{\boldsymbol{\Theta}}$)**.** *For any $\alpha > 0$ and all $\boldsymbol{f} : \boldsymbol{\Theta} \to \mathbb{R}^{m+n}$, there exist positive constants c_α and C_α, depending only on α, such that*

$$c_\alpha \|\boldsymbol{f}\|_{\boldsymbol{\Theta}} \leqslant \|\boldsymbol{f}\|_{\boldsymbol{\Theta},\alpha} \leqslant C_\alpha \|\boldsymbol{f}\|_{\boldsymbol{\Theta}} \tag{B.1}$$

Proof of Lemma B.1. For a fixed $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, write $\boldsymbol{f}(\boldsymbol{\theta}) = \big(\boldsymbol{u}(\boldsymbol{\theta})^\top, \boldsymbol{v}(\boldsymbol{\theta})^\top\big)^\top$. Recall $\|\boldsymbol{f}(\boldsymbol{\theta})\|_\alpha = \|\boldsymbol{u}(\boldsymbol{\theta})\| + \alpha \|\boldsymbol{v}(\boldsymbol{\theta})\|$. We have

$$\begin{aligned}\|\boldsymbol{f}(\boldsymbol{\theta})\|_\alpha &\geqslant \min\{1,\alpha\}\big(\|\boldsymbol{u}(\boldsymbol{\theta})\| + \|\boldsymbol{v}(\boldsymbol{\theta})\|\big)\\ &\geqslant \min\{1,\alpha\}\sqrt{\|\boldsymbol{u}(\boldsymbol{\theta})\|^2 + \|\boldsymbol{v}(\boldsymbol{\theta})\|^2} = c_\alpha \|\boldsymbol{f}(\boldsymbol{\theta})\|,\end{aligned} \tag{B.2}$$

where $c_\alpha := \min\{1,\alpha\}$. Second, using $a + b \leqslant \sqrt{2(a^2+b^2)}$ for $a, b \geqslant 0$, we have

$$\begin{aligned}\|\boldsymbol{f}(\boldsymbol{\theta})\|_\alpha &\leqslant \max\{1,\alpha\}\big(\|\boldsymbol{u}(\boldsymbol{\theta})\| + \|\boldsymbol{v}(\boldsymbol{\theta})\|\big)\\ &\leqslant \sqrt{2}\max\{1,\alpha\}\sqrt{\|\boldsymbol{u}(\boldsymbol{\theta})\|^2 + \|\boldsymbol{v}(\boldsymbol{\theta})\|^2} = C_\alpha \|\boldsymbol{f}(\boldsymbol{\theta})\|,\end{aligned} \tag{B.3}$$

where $C_\alpha := \sqrt{2}\max\{1,\alpha\}$. Taking $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}$ on both sides yields Lemma B.1. □

Note that Lemma B.1 implies that, for a sequence of functions $\{\boldsymbol{f}_t\}$, we have $\|\boldsymbol{f}_t\|_{\boldsymbol{\Theta},\alpha} \to 0$ if and only if $\|\boldsymbol{f}_t\|_{\boldsymbol{\Theta}} \to 0$ as $t \to \infty$.

Lemma B.2. *Consider a block matrix $\boldsymbol{M} = \left(\begin{smallmatrix}\boldsymbol{M}_{11} & \boldsymbol{M}_{12}\\ \boldsymbol{M}_{21} & \boldsymbol{M}_{22}\end{smallmatrix}\right) \in \mathbb{R}^{p\times(m+n)}$ for $p, m, n \in \mathbb{Z}^+$. Then*

$$\|\boldsymbol{M}\|_\alpha \leqslant \max\big\{\|\boldsymbol{M}_{11}\| + \alpha\|\boldsymbol{M}_{21}\|,\ \|\boldsymbol{M}_{22}\| + \alpha^{-1}\|\boldsymbol{M}_{12}\|\big\}, \qquad \forall \alpha > 0. \tag{B.4}$$

Proof of Lemma B.2. Let $\boldsymbol{x} = (\boldsymbol{x}_1^\top, \boldsymbol{x}_2^\top)^\top \neq \boldsymbol{0}$ with $\boldsymbol{x}_1 \in \mathbb{R}^m$ and $\boldsymbol{x}_2 \in \mathbb{R}^n$. Note that $\boldsymbol{M}\boldsymbol{x} = \left(\begin{smallmatrix}\boldsymbol{M}_{11}\boldsymbol{x}_1+\boldsymbol{M}_{12}\boldsymbol{x}_2\\ \boldsymbol{M}_{21}\boldsymbol{x}_1+\boldsymbol{M}_{22}\boldsymbol{x}_2\end{smallmatrix}\right)$. By definition of $\|\cdot\|_\alpha$ in (A.6) and the triangle inequality,

$$\begin{aligned}
\|\boldsymbol{M}\boldsymbol{x}\|_\alpha &= \|\boldsymbol{M}_{11}\boldsymbol{x}_1 + \boldsymbol{M}_{12}\boldsymbol{x}_2\| + \alpha\,\|\boldsymbol{M}_{21}\boldsymbol{x}_1 + \boldsymbol{M}_{22}\boldsymbol{x}_2\| \\
&\leqslant \big(\,\|\boldsymbol{M}_{11}\| + \alpha\,\|\boldsymbol{M}_{21}\|\,\big)\,\|\boldsymbol{x}_1\| + \big(\,\|\boldsymbol{M}_{12}\| + \alpha\,\|\boldsymbol{M}_{22}\|\,\big)\,\|\boldsymbol{x}_2\| \\
&= \big(\,\|\boldsymbol{M}_{11}\| + \alpha\,\|\boldsymbol{M}_{21}\|\,\big)\,\|\boldsymbol{x}_1\| + \big(\,\|\boldsymbol{M}_{22}\| + \alpha^{-1}\,\|\boldsymbol{M}_{12}\|\,\big)\,\alpha\,\|\boldsymbol{x}_2\| \\
&\leqslant \max\big\{\,\|\boldsymbol{M}_{11}\| + \alpha\,\|\boldsymbol{M}_{21}\|\,,\ \|\boldsymbol{M}_{22}\| + \alpha^{-1}\,\|\boldsymbol{M}_{12}\|\,\big\}\big(\,\|\boldsymbol{x}_1\| + \alpha\,\|\boldsymbol{x}_2\|\,\big).
\end{aligned}$$

This implies $\|\boldsymbol{M}\boldsymbol{x}\|_\alpha/\|\boldsymbol{x}\|_\alpha \leqslant \max\{\|\boldsymbol{M}_{11}\| + \alpha\,\|\boldsymbol{M}_{21}\|\,,\ \|\boldsymbol{M}_{22}\| + \alpha^{-1}\,\|\boldsymbol{M}_{12}\|\}$ for all $\boldsymbol{x} \neq \boldsymbol{0}$. Lemma B.2 is then obtained by taking the supremum over $\boldsymbol{x}$. □

Lemma B.3. *Suppose that the filter dynamics are generated by the Gaussian quasi-likelihood recursion in (2.3) and Proposition 1 under the true parameter value $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$. Suppose further that for every $t \in \mathbb{Z}$ and for almost every realization of $\mathscr{F}_{t-1}$, the conditional distribution of $\boldsymbol{y}_t \mid \mathscr{F}_{t-1}$ admits a pdf with a support that contains a nonempty open subset of $\mathbb{R}^k$. Let $\boldsymbol{\kappa} : \mathbb{Z}^{k-1} \times \mathbb{Z}^{k-1} \to \mathbb{R}^{k-1}$, $\boldsymbol{G}, \boldsymbol{H} : \mathbb{Z}^{k-1} \times \mathbb{Z}^{k-1} \to \mathbb{R}^{(k-1)\times(k-1)}$ be nonrandom maps.*

(i) For each $t \in \mathbb{Z}$, if $\det(\boldsymbol{A}_{\lambda 0}) \neq 0$, then $\boldsymbol{\lambda}_t(\boldsymbol{\theta}_0)$ admits a pdf with a support containing a nonempty open subset of $\mathbb{R}^k$. Moreover, for constant coefficient matrices $\boldsymbol{\Delta}_{\omega_\lambda}$, $\boldsymbol{\Delta}_{B_\lambda}$, and $\boldsymbol{\Delta}_{A_\lambda}$, if

$$\boldsymbol{\Delta}_{\omega_\lambda} + \boldsymbol{\Delta}_{B_\lambda}\,\boldsymbol{\lambda}_t(\boldsymbol{\theta}_0) + \boldsymbol{\Delta}_{A_\lambda}\left(\boldsymbol{V}_t(\boldsymbol{\theta}_0)^\top \boldsymbol{y}_t\right)^{\odot 2} = \boldsymbol{0} \quad a.s., \tag{B.5}$$

then $(\boldsymbol{\Delta}_{\omega_\lambda}, \boldsymbol{\Delta}_{B_\lambda}, \boldsymbol{\Delta}_{A_\lambda}) = (\boldsymbol{0}, \boldsymbol{0}, \boldsymbol{0})$.

(ii) For each $t \in \mathbb{Z}$, if $\det(\boldsymbol{A}_{\varphi 0}) \neq 0$, then $\boldsymbol{\varphi}_t(\boldsymbol{\theta}_0)$ admits a pdf with a support that contains a nonempty open subset of $\mathbb{R}^{k-1}$. Moreover, for each $t \in \mathbb{Z}$ and almost surely, the conditional distribution of $\partial \ell_t(\boldsymbol{f}_t, \boldsymbol{\nu})/\partial \boldsymbol{\varphi}_t\big|_{\boldsymbol{\varphi}_t = \boldsymbol{\varphi}_t(\boldsymbol{\theta}_0),\, \boldsymbol{\lambda}_t = \boldsymbol{\lambda}_t(\boldsymbol{\theta}_0)}$ given $\mathscr{F}_{t-1}$ is absolutely continuous with respect to Lebesgue measure with a support that contains

a nonempty open subset of $\mathbb{R}^{k-1}$*. If*

$$\boldsymbol{\kappa}(\boldsymbol{q},\boldsymbol{q}') + \boldsymbol{G}(\boldsymbol{q},\boldsymbol{q}')\,\boldsymbol{\varphi}_t(\boldsymbol{\theta}_0) + \boldsymbol{H}(\boldsymbol{q},\boldsymbol{q}')\,\frac{\partial \ell_t(\boldsymbol{f}_t,\boldsymbol{\nu})}{\partial \boldsymbol{\varphi}_t}\bigg|_{\boldsymbol{\varphi}_t=\boldsymbol{\varphi}_t(\boldsymbol{\theta}_0),\ \boldsymbol{\lambda}_t=\boldsymbol{\lambda}_t(\boldsymbol{\theta}_0)} = \mathbf{0}, \quad a.s. \tag{B.6}$$

for a nonrandom pair $(\boldsymbol{q},\boldsymbol{q}') \in \mathbb{Z}^{k-1}\times\mathbb{Z}^{k-1}$*, then* $\boldsymbol{\kappa}(\boldsymbol{q},\boldsymbol{q}') = \mathbf{0}$*,* $\boldsymbol{G}(\boldsymbol{q},\boldsymbol{q}') = \mathbf{0}$*, and* $\boldsymbol{H}(\boldsymbol{q},\boldsymbol{q}') = \mathbf{0}$*.*

(iii) For each $t \in \mathbb{Z}$*, the event* E_t *defined in* (A.13) *occurs almost surely.*

Proof of Lemma B.3*. Proof of Lemma B.3(i):* Fix $t \in \mathbb{Z}$. Conditional on $\mathscr{F}_{t-1}$, both $\boldsymbol{\Lambda}_t(\boldsymbol{\theta}_0)$ and $\boldsymbol{V}_t(\boldsymbol{\theta}_0)$ are fixed. Since $\boldsymbol{V}_t(\boldsymbol{\theta}_0)$ is orthogonal (and thus invertible), the change-of-variables formula implies that, almost surely, the conditional distribution of $\boldsymbol{u}_t := \boldsymbol{V}_t(\boldsymbol{\theta}_0)^\top \boldsymbol{y}_t$ given $\mathscr{F}_{t-1}$ also admits a density that is positive on a nonempty open subset of $\mathbb{R}^k$. Let $\mathbb{U}_t$ be a nonempty open set on which the conditional density of $\boldsymbol{u}_t$ given $\mathscr{F}_{t-1}$ is positive and all coordinates of $\boldsymbol{u}_t$ are nonzero. Then $\{\boldsymbol{u}^{\odot 2} : \boldsymbol{u} \in \mathbb{U}_t\}$ is a nonempty open subset of $\mathbb{R}^k$. Therefore, conditional on $\mathscr{F}_{t-1}$, the random vector $\boldsymbol{u}_t^{\odot 2}$ has support containing a nonempty open subset of $\mathbb{R}^k$. Under the true eigenvalue recursion, we have $\boldsymbol{\lambda}_{t+1}(\boldsymbol{\theta}_0) = (\boldsymbol{I}_k - \boldsymbol{B}_{\lambda 0})\boldsymbol{\omega}_{\lambda 0} + (\boldsymbol{B}_{\lambda 0} - \boldsymbol{A}_{\lambda 0})\,\boldsymbol{\lambda}_t(\boldsymbol{\theta}_0) + \boldsymbol{A}_{\lambda 0}\,\boldsymbol{u}_t^{\odot 2}$. Since $\det(\boldsymbol{A}_{\lambda 0}) \neq 0$ by assumption, the map $\boldsymbol{\xi} \mapsto (\boldsymbol{I}_k - \boldsymbol{B}_{\lambda 0})\boldsymbol{\omega}_{\lambda 0} + (\boldsymbol{B}_{\lambda 0} - \boldsymbol{A}_{\lambda 0})\boldsymbol{\lambda}_t(\boldsymbol{\theta}_0) + \boldsymbol{A}_{\lambda 0}\,\boldsymbol{\xi}$ has nonsingular Jacobian $\boldsymbol{A}_{\lambda 0}$ with respect to $\boldsymbol{\xi}$, and therefore maps nonempty open sets in $\mathbb{R}^k$ to nonempty open sets by the inverse function theorem of Rudin (1976, Theorem 9.24 (a)). It follows that, for almost every realization of $\mathscr{F}_{t-1}$, the conditional support of $\boldsymbol{\lambda}_{t+1}(\boldsymbol{\theta}_0)$ given $\mathscr{F}_{t-1}$ contains a nonempty open subset of $\mathbb{R}^k$.

To continue, suppose that (B.5) holds. Note that $\mathbb{P}(G) = \mathbb{E}\big[\mathbb{E}\big(\mathbb{1}_G \,|\, \mathscr{H}\big)\big]$ for a σ-field $\mathscr{H}$ and any event G, where $\mathbb{1}_{\{\cdot\}}$ is an indicator function. Let $G^\complement$ be the complement of G. Setting $G = \big\{\boldsymbol{\Delta}_{\omega_\lambda} + \boldsymbol{\Delta}_{B_\lambda}\,\boldsymbol{\lambda}_t(\boldsymbol{\theta}_0) + \boldsymbol{\Delta}_{A_\lambda}\,\big(\boldsymbol{V}_t(\boldsymbol{\theta}_0)^\top \boldsymbol{y}_t\big)^{\odot 2} = \mathbf{0}\big\}$, and $\mathscr{H} = \mathscr{F}_{t-1}$, we have $0 = \mathbb{P}(G^\complement) = \mathbb{E}\big[\mathbb{E}\big(\mathbb{1}_{G^\complement} \,|\, \mathscr{H}\big)\big] \geqslant 0$, and thus $\mathbb{E}\big(\mathbb{1}_{G^\complement} \,|\, \mathscr{H}\big) = \mathbb{P}(G^\complement \mid \mathscr{H}) = 0$ a.s. Equivalently, $\mathbb{P}(G \mid \mathscr{H}) = 1$ a.s. That is, (B.5) also holds with conditional probability one given $\mathscr{F}_{t-1}$ a.s. For almost every realization of

$\mathscr{F}_{t-1}$, $\boldsymbol{\lambda}_t(\boldsymbol{\theta}_0)$ and $\boldsymbol{V}_t(\boldsymbol{\theta}_0)$ are fixed, and, by the argument above, $\boldsymbol{u}_t^{\odot 2}$ has conditional support containing a nonempty open subset of $\mathbb{R}^k$. It follows that the map $\boldsymbol{\xi} \mapsto \boldsymbol{\Delta}_{\omega_\lambda} + \boldsymbol{\Delta}_{B_\lambda} \boldsymbol{\lambda}_t(\boldsymbol{\theta}_0) + \boldsymbol{\Delta}_{A_\lambda} \boldsymbol{\xi}$ is equal to zero on a nonempty open subset of $\mathbb{R}^k$. This is possible only if $\boldsymbol{\Delta}_{A_\lambda} = \mathbf{0}$. The equality therefore reduces to $\boldsymbol{\Delta}_{\omega_\lambda} + \boldsymbol{\Delta}_{B_\lambda} \boldsymbol{\lambda}_t(\boldsymbol{\theta}_0) = \mathbf{0}$ a.s. Since $\boldsymbol{\lambda}_t(\boldsymbol{\theta}_0)$ conditional on $\mathscr{F}_{t-2}$ has support containing a nonempty open subset of $\mathbb{R}^k$ and using a similar argument as before, the map $\boldsymbol{\xi} \mapsto \boldsymbol{\Delta}_{\omega_\lambda} + \boldsymbol{\Delta}_{B_\lambda} \boldsymbol{\xi}$ must be zero on a nonempty open subset of $\mathbb{R}^k$. It follows that $\boldsymbol{\Delta}_{B_\lambda} = \mathbf{0}$ and $\boldsymbol{\Delta}_{\omega_\lambda} = \mathbf{0}$. Hence $(\boldsymbol{\Delta}_{\omega_\lambda}, \boldsymbol{\Delta}_{B_\lambda}, \boldsymbol{\Delta}_{A_\lambda}) = (\mathbf{0}, \mathbf{0}, \mathbf{0})$.

Proof of Lemma B.3(ii): Note that the preceding argument implies that the conditional distribution of $\boldsymbol{\lambda}_t(\boldsymbol{\theta}_0)$ given $\mathscr{F}_{t-2}$ is absolutely continuous. Its unconditional distribution is therefore also absolutely continuous, using the identity $\mathbb{P}(G) = \mathbb{E}\big[\mathbb{P}(G \mid \mathscr{H})\big]$ for any event G and σ-field $\mathscr{H}$. Namely, for each fixed $t \in \mathbb{Z}$, the entries of $\boldsymbol{\lambda}_t(\boldsymbol{\theta}_0)$ are pairwise distinct a.s., that is,

$$\mathbb{P}\Big(\lambda_{i,t}(\boldsymbol{\theta}_0) = \lambda_{j,t}(\boldsymbol{\theta}_0)\Big) = 0, \qquad i \neq j. \tag{B.7}$$

We shall work on this almost sure event. Define $\boldsymbol{x}_t := \boldsymbol{\Lambda}_t(\boldsymbol{\theta}_0)^{-1/2} \boldsymbol{V}_t(\boldsymbol{\theta}_0)^\top \boldsymbol{y}_t$. Since $\boldsymbol{\Lambda}_t(\boldsymbol{\theta}_0)^{-1/2} \boldsymbol{V}_t(\boldsymbol{\theta}_0)^\top$ is nonsingular almost surely, $\boldsymbol{x}_t \mid \mathscr{F}_{t-1}$ admits a density that is positive on a nonempty open subset of $\mathbb{R}^k$ almost surely. For fixed $\boldsymbol{\lambda}_t(\boldsymbol{\theta}_0)$ and $\boldsymbol{\varphi}_t(\boldsymbol{\theta}_0)$, define the map $\boldsymbol{g}_t : \mathbb{R}^k \to \mathbb{R}^{k-1}$ by

$$\boldsymbol{g}_t(\boldsymbol{x}) := -\begin{pmatrix} \boldsymbol{x}^\top \boldsymbol{\Lambda}_t(\boldsymbol{\theta}_0)^{1/2} \boldsymbol{S}(\boldsymbol{\varphi}_t(\boldsymbol{\theta}_0))^\top [\mathscr{D}_2 \boldsymbol{S}(\boldsymbol{\varphi}_t(\boldsymbol{\theta}_0))] \boldsymbol{\Lambda}_t(\boldsymbol{\theta}_0)^{-1/2} \boldsymbol{x} \\ \vdots \\ \boldsymbol{x}^\top \boldsymbol{\Lambda}_t(\boldsymbol{\theta}_0)^{1/2} \boldsymbol{S}(\boldsymbol{\varphi}_t(\boldsymbol{\theta}_0))^\top [\mathscr{D}_k \boldsymbol{S}(\boldsymbol{\varphi}_t(\boldsymbol{\theta}_0))] \boldsymbol{\Lambda}_t(\boldsymbol{\theta}_0)^{-1/2} \boldsymbol{x} \end{pmatrix}.$$

Then $\boldsymbol{g}_t(\boldsymbol{x}_t)$ is exactly the true Gaussian rotation angle score entering the recursion for $\boldsymbol{\varphi}_t(\boldsymbol{\theta}_0)$.

To apply the inverse function theorem (see, e.g., Rudin, 1976, Theorem 9.24), we now verify that the Jacobian of $\boldsymbol{g}_t$ has full row rank $k-1$ on a nonempty open subset of

$\mathbb{R}^k$. We fix the first coordinate of $\boldsymbol{x} = (x_1, x_2, \ldots, x_k)^\top$ and use the $(k-1) \times (k-1)$ submatrix of the Jacobian formed by columns $2, \ldots, k$. Nonsingularity of this submatrix implies that $\boldsymbol{g}_t$ maps local variation in $(x_2, \ldots, x_k)^\top$ into a nonempty open subset of $\mathbb{R}^{k-1}$ due to the inverse function theorem.

To establish the nonsingularity of the Jacobian, let $\widetilde{\boldsymbol{S}}_{i,t} := \boldsymbol{S}(\boldsymbol{\varphi}_t(\boldsymbol{\theta}_0))^\top[\mathscr{D}_i \boldsymbol{S}(\boldsymbol{\varphi}_t(\boldsymbol{\theta}_0))]$ for $i = 2, \ldots, k$, and note that $\widetilde{\boldsymbol{S}}_{i,t}^\top = -\widetilde{\boldsymbol{S}}_{i,t}$. The latter follows directly from Eq. (B.8) below. For $i = 2, \ldots, k$, the Jacobian of the ith component $g_{i,t}$ of $\boldsymbol{g}_t$ is then given by

$$\frac{\partial g_{i,t}(\boldsymbol{x})}{\partial \boldsymbol{x}^\top} = -\boldsymbol{x}^\top \Big(\boldsymbol{\Lambda}_t(\boldsymbol{\theta}_0)^{1/2} \widetilde{\boldsymbol{S}}_{i,t} \boldsymbol{\Lambda}_t(\boldsymbol{\theta}_0)^{-1/2} - \boldsymbol{\Lambda}_t(\boldsymbol{\theta}_0)^{-1/2} \widetilde{\boldsymbol{S}}_{i,t} \boldsymbol{\Lambda}_t(\boldsymbol{\theta}_0)^{1/2} \Big).$$

Furthermore, we observe that

$$\widetilde{\boldsymbol{S}}_{i,t} = \left(\prod_{\ell=i+1}^{k} \boldsymbol{R}_{\ell,1}\big(\varphi_{\ell,t}(\boldsymbol{\theta}_0)\big) \right)^\top \big(\boldsymbol{e}_1 \boldsymbol{e}_i^\top - \boldsymbol{e}_i \boldsymbol{e}_1^\top\big) \left(\prod_{\ell=i+1}^{k} \boldsymbol{R}_{\ell,1}\big(\varphi_{\ell,t}(\boldsymbol{\theta}_0)\big) \right), \tag{B.8}$$

where $\boldsymbol{e}_i$ denotes the ith canonical basis vector of $\mathbb{R}^k$, with 1 in the ith entry and 0 elsewhere. Since $\prod_{\ell=i+1}^{k} \boldsymbol{R}_{\ell,1}\big(\varphi_{\ell,t}(\boldsymbol{\theta}_0)\big)$ leaves the ith coordinate unchanged, this matrix has the form $\widetilde{\boldsymbol{S}}_{i,t} = \boldsymbol{\eta}_{i,t} \boldsymbol{e}_i^\top - \boldsymbol{e}_i \boldsymbol{\eta}_{i,t}^\top$, for the unit length vector $\boldsymbol{\eta}_{i,t} := \Big(\prod_{\ell=i+1}^{k} \boldsymbol{R}_{\ell,1}\big(\varphi_{\ell,t}(\boldsymbol{\theta}_0)\big) \Big)^\top \boldsymbol{e}_1 \in \operatorname{span}\{\boldsymbol{e}_1, \boldsymbol{e}_{i+1}, \ldots, \boldsymbol{e}_k\}$ with $\|\boldsymbol{\eta}_{i,t}\| = 1$. Therefore, among columns $2, \ldots, k$, the row vector $\partial g_{i,t}(\boldsymbol{x})/\partial \boldsymbol{x}^\top$ can have nonzero entries only in columns $i, i+1, \ldots, k$ for $i \geqslant 2$, i.e., it has an upper-triangular structure. With regard to its diagonal entries, note that the ith diagonal entry for $i \geqslant 2$ is given by

$$\frac{\partial g_{i,t}(\boldsymbol{x})}{\partial \boldsymbol{x}^\top} \boldsymbol{e}_i = -\boldsymbol{x}^\top \Big(\boldsymbol{\Lambda}_t(\boldsymbol{\theta}_0)^{1/2} \lambda_{i,t}(\boldsymbol{\theta}_0)^{-1/2} - \boldsymbol{\Lambda}_t(\boldsymbol{\theta}_0)^{-1/2} \lambda_{i,t}(\boldsymbol{\theta}_0)^{1/2} \Big) \boldsymbol{\eta}_{i,t},$$

where we used $\boldsymbol{\eta}_{i,t} \in \operatorname{span}\{\boldsymbol{e}_1, \boldsymbol{e}_{i+1}, \ldots, \boldsymbol{e}_k\}$ and thus $\boldsymbol{\eta}_{i,t}^\top \boldsymbol{e}_i = 0$. Since $\boldsymbol{\eta}_{i,t}$ is a unit length vector with zero ith coordinate, it has at least one nonzero coordinate outside the ith position. Since the entries of $\boldsymbol{\lambda}_t(\boldsymbol{\theta}_0)$ are pairwise distinct a.s. by (B.7), for each $i = 2, \ldots, k$, the sets $\big\{\boldsymbol{x} \in \mathbb{R}^k : \big(\partial g_{i,t}(\boldsymbol{x})/\partial \boldsymbol{x}^\top\big) \boldsymbol{e}_i = 0\big\}$ all have Lebesgue measure zero.

Since $\boldsymbol{x}_t \mid \mathscr{F}_{t-1}$ admits a pdf, the diagonal elements of the above Jacobian submatrix are strictly non-zero almost surely. Viewing $\boldsymbol{g}_t$ as a map from the remaining $k-1$ coordinates to $\mathbb{R}^{k-1}$ after fixing the first coordinate of $\boldsymbol{x}$ (at its realized value), the change-of-variables formula implies that $\boldsymbol{g}_t(\boldsymbol{x}_t) \mid \mathscr{F}_{t-1}$ also admits a pdf on $\mathbb{R}^{k-1}$. Moreover, since the diagonal elements of the upper triangular Jacobian submatrix are continuous functions of $\boldsymbol{x}$, their nonzero values at $\boldsymbol{x}_t$ remain nonzero in a neighborhood of $\boldsymbol{x}_t$ almost surely. Thus, after fixing the first coordinate of $\boldsymbol{x}$, the inverse function theorem applied to the remaining coordinates of $\boldsymbol{x}$ implies that the image of a local open subset under $\boldsymbol{g}_t$ contains a nonempty open subset of $\mathbb{R}^{k-1}$. Hence the support of $\boldsymbol{g}_t(\boldsymbol{x}_t) \mid \mathscr{F}_{t-1}$ contains a nonempty open subset of $\mathbb{R}^{k-1}$. Since the true recursion can be written as $\boldsymbol{\varphi}_{t+1}(\boldsymbol{\theta}_0) = (\boldsymbol{I}_{k-1} - \boldsymbol{B}_{\varphi 0})\,\boldsymbol{\omega}_{\varphi 0} + \boldsymbol{B}_{\varphi 0}\,\boldsymbol{\varphi}_t(\boldsymbol{\theta}_0) + \boldsymbol{A}_{\varphi 0}\,\boldsymbol{g}_t(\boldsymbol{x}_t)$ with $\det(\boldsymbol{A}_{\varphi 0}) \neq 0$ by assumption, and since $\boldsymbol{g}_t(\boldsymbol{x}_t) \mid \mathscr{F}_{t-1}$ admits a pdf with conditional support containing a nonempty open subset of $\mathbb{R}^{k-1}$, the conditional distribution of $\boldsymbol{\varphi}_{t+1}(\boldsymbol{\theta}_0)$ given $\mathscr{F}_{t-1}$ also admits a pdf, and its conditional support contains a nonempty open subset of $\mathbb{R}^{k-1}$. Consequently, the unconditional distribution of $\boldsymbol{\varphi}_{t+1}(\boldsymbol{\theta}_0)$ also admits a pdf on $\mathbb{R}^{k-1}$.

Finally, we consider (B.6). For each fixed pair $(\boldsymbol{q}, \boldsymbol{q}') \in \mathbb{Z}^{k-1} \times \mathbb{Z}^{k-1}$, conditional on $\mathscr{F}_{t-1}$, $\boldsymbol{\varphi}_t(\boldsymbol{\theta}_0)$ is fixed, while the conditional support of $\boldsymbol{g}_t(\boldsymbol{x}_t)$ contains a nonempty open subset of $\mathbb{R}^{k-1}$. Hence, by similar arguments as for Part (i) and given the above results, we obtain $\boldsymbol{H}(\boldsymbol{q}, \boldsymbol{q}') = \boldsymbol{0}$. Eq. (B.6) then reduces to $\boldsymbol{\kappa}(\boldsymbol{q}, \boldsymbol{q}') + \boldsymbol{G}(\boldsymbol{q}, \boldsymbol{q}')\boldsymbol{\varphi}_t(\boldsymbol{\theta}_0) = \boldsymbol{0}$ a.s. Again, by a similar argument as for Part (i), since the support of $\boldsymbol{\varphi}_t(\boldsymbol{\theta}_0) \mid \mathscr{F}_{t-2}$ also contains a nonempty open subset, we get that $\boldsymbol{G}(\boldsymbol{q}, \boldsymbol{q}') = \boldsymbol{0}$, and thus, finally, that $\boldsymbol{\kappa}(\boldsymbol{q}, \boldsymbol{q}') = \boldsymbol{0}$.

Proof of Lemma B.3(iii): Since $\boldsymbol{\varphi}_t(\boldsymbol{\theta}_0)$ admits a pdf on $\mathbb{R}^{k-1}$, it assigns zero probability to any set of Lebesgue measure zero in $\mathbb{R}^{k-1}$. We have $\mathbb{P}\big(\cos\big(\varphi_{i,t}(\boldsymbol{\theta}_0)\big) \neq 0\big) = 1$ for $i = 2, \ldots, k$, and $\mathbb{P}\big(\sin\big(\varphi_{k,t}(\boldsymbol{\theta}_0)\big) \neq 0\big) = 1$. Together with (B.7), this yields $\mathbb{P}(E_t) = 1$ for each fixed $t \in \mathbb{Z}$. $\square$

Lemma B.4 (Column-structure uniqueness)**.** *For $k \geqslant 2$, recall that $\boldsymbol{S}(\boldsymbol{\varphi}) = \prod_{i=2}^{k} \boldsymbol{R}_{i,1}(\varphi_i)$, where $\boldsymbol{\varphi} = (\varphi_2, \ldots, \varphi_k)^\top$. Suppose that $\cos(\varphi_i) \neq 0$ for $i = 2, \ldots, k$ and $\sin(\varphi_k) \neq 0$. If there exist a permutation matrix $\boldsymbol{\Pi}$, a signed matrix $\boldsymbol{D} = \mathrm{diag}(\pm 1, \pm 1, \ldots, \pm 1)$, and a vector $\tilde{\boldsymbol{\varphi}}$ such that $\boldsymbol{S}(\tilde{\boldsymbol{\varphi}}) = \boldsymbol{S}(\boldsymbol{\varphi})\boldsymbol{D}\boldsymbol{\Pi}$, then $\boldsymbol{\Pi} \in \{\boldsymbol{I}_k, \boldsymbol{\Pi}_{1k}\}$, where $\boldsymbol{\Pi}_{1k} = \big(\boldsymbol{e}_k, \boldsymbol{e}_2, \boldsymbol{e}_3, \ldots, \boldsymbol{e}_{k-1}, \boldsymbol{e}_1\big)$, and $\boldsymbol{e}_i$ denotes the ith canonical basis vector in $\mathbb{R}^k$.*

Proof of Lemma B.4. Note that when $k = 2$, any permutation matrix $\boldsymbol{\Pi}$ necessarily satisfies $\boldsymbol{\Pi} \in \{\boldsymbol{I}_k, \boldsymbol{\Pi}_{1k}\}$, so the above result holds trivially.

In what follows, let $k \geqslant 3$. For any matrices $\boldsymbol{U} = (\boldsymbol{U}_{\cdot,1}, \ldots, \boldsymbol{U}_{\cdot,k})$ and $\boldsymbol{V} = (\boldsymbol{V}_{\cdot,1}, \ldots, \boldsymbol{V}_{\cdot,k})$, where $\boldsymbol{U}_{\cdot,i}$ and $\boldsymbol{V}_{\cdot,i}$ denote the ith columns for $i = 1, \ldots, k$, if $\boldsymbol{U} = \boldsymbol{V}\boldsymbol{R}_{i,1}(\varphi)$, then we have

$$\boldsymbol{U}_{\cdot,1} = \cos(\varphi)\boldsymbol{V}_{\cdot,1} - \sin(\varphi)\boldsymbol{V}_{\cdot,i}, \quad \boldsymbol{U}_{\cdot,i} = \sin(\varphi)\boldsymbol{V}_{\cdot,1} + \cos(\varphi)\boldsymbol{V}_{\cdot,i}, \quad \boldsymbol{U}_{\cdot,\ell} = \boldsymbol{V}_{\cdot,\ell}, \quad \ell \notin \{1, i\}.$$

Namely, right-multiplying $\boldsymbol{V}$ by $\boldsymbol{R}_{i,1}(\cdot)$ only changes the first and the ith columns. Define $\boldsymbol{V}^{(i)} = \boldsymbol{V}^{(i-1)}\boldsymbol{R}_{i,1}(\varphi_i)$ for $i = 2, \ldots, k$, with $\boldsymbol{V}^{(1)} = \boldsymbol{I}_k = (\boldsymbol{e}_1, \ldots, \boldsymbol{e}_k)$, where $\boldsymbol{e}_j$ denotes the jth standard basis vector of $\mathbb{R}^k$. Clearly, $\boldsymbol{S}(\boldsymbol{\varphi}) = \boldsymbol{V}^{(k)}$.

Fix any $j = 2, 3, \ldots, k$ and note that:

(a) for $i < j$, since each step only modifies the first and the ith columns, the jth column remains unchanged; that is, $\boldsymbol{V}^{(i)}_{\cdot,j} = \boldsymbol{e}_j$.

(b) for $i = j$,

$$\boldsymbol{V}^{(j)}_{\cdot,j} = \sin(\varphi_j)\boldsymbol{V}^{(j-1)}_{\cdot,1} + \cos(\varphi_j)\boldsymbol{V}^{(j-1)}_{\cdot,j} = \sin(\varphi_j)\boldsymbol{V}^{(j-1)}_{\cdot,1} + \cos(\varphi_j)\boldsymbol{e}_j. \tag{B.9}$$

Over the first $j-1$ steps, the first column $\boldsymbol{V}^{(j-1)}_{\cdot,1}$ only interacts with columns $2, \ldots, j-1$. As a result, the $j, j+1, \ldots, k$ rows of $\boldsymbol{V}^{(j-1)}_{\cdot,1}$ are all zero. This, in turn, implies that the $j+1, j+2, \ldots, k$ rows of $\boldsymbol{V}^{(j)}_{\cdot,j}$ are zero, provided that $\cos(\varphi_j) \neq 0$.

(c) for $i > j$, the column $\boldsymbol{V}^{(i)}_{\cdot,j}$ remains unchanged and satisfies $\boldsymbol{V}^{(i)}_{\cdot,j} = \boldsymbol{V}^{(j)}_{\cdot,j}$. In words, for any column $j \geqslant 2$ of $\boldsymbol{S}(\boldsymbol{\varphi}) = \boldsymbol{V}^{(k)}$ with $k \geqslant 3$, all entries in rows $j+1, \ldots, k$ for column j are equal to zero, and the largest row index at which a nonzero entry can occur is j in column j.

Note that for any $\boldsymbol{\varphi}$ and $\tilde{\boldsymbol{\varphi}}$, $\boldsymbol{S}(\boldsymbol{\varphi})\boldsymbol{D}$ and $\boldsymbol{S}(\tilde{\boldsymbol{\varphi}})$ must share the same column zero structure described above. Since right multiplication of $\boldsymbol{S}(\boldsymbol{\varphi})\boldsymbol{D}$ by $\boldsymbol{\Pi}$ merely permutes its columns, if $\boldsymbol{S}(\tilde{\boldsymbol{\varphi}}) = \boldsymbol{S}(\boldsymbol{\varphi})\boldsymbol{D}\boldsymbol{\Pi}$, then each column of $\boldsymbol{S}(\boldsymbol{\varphi})\boldsymbol{D}$ must be mapped to a column position with an identical zero structure, which implies that the columns $2, 3, \ldots, k-1$ of $\boldsymbol{S}(\boldsymbol{\varphi})\boldsymbol{D}$ must be fixed. However, the first and last columns may exhibit the same nonzero structure and may therefore be interchanged, which implies that $\boldsymbol{\Pi} \in \{\boldsymbol{I}_k, \boldsymbol{\Pi}_{1k}\}$. This completes the proof. $\square$

Lemma B.5. *Recall* $\boldsymbol{S}(\boldsymbol{\varphi})$ *from* (2.1)*, where* $\boldsymbol{\varphi} = (\varphi_2, \ldots, \varphi_k)^\top$*. For any integer vector* $\boldsymbol{q} = (q_2, \ldots, q_k)^\top \in \mathbb{Z}^{k-1}$*, define the* $k \times k$ *diagonal sign matrix* $\boldsymbol{D}(\boldsymbol{q}) := \prod_{i=2}^{k} \boldsymbol{D}_{1i}^{q_i}$*, where* $\boldsymbol{D}_{1i} = (-\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_{i-1}, -\boldsymbol{e}_i, \boldsymbol{e}_{i+1}, \ldots, \boldsymbol{e}_k)$ *for canonical basis vectors* $\boldsymbol{e}_i$*, or in words,* $\boldsymbol{D}_{1i}$ *is an identity matrix with the* $(1,1)$ *and* (i,i) *diagonal entries replaced by* -1*. Also define the diagonal matrix* $\boldsymbol{\zeta}(\boldsymbol{q}) := \operatorname{diag}\big(\zeta_2(\boldsymbol{q}), \ldots, \zeta_k(\boldsymbol{q})\big)$ *acting componentwise on* $\boldsymbol{\varphi}$*, where*

$$\zeta_2(\boldsymbol{q}) := 1, \qquad \zeta_i(\boldsymbol{q}) := (-1)^{\sum_{j=2}^{i-1} q_j}, \qquad i = 3, \ldots, k.$$

Then, for every $\boldsymbol{\varphi} \in \mathbb{R}^{k-1}$*, we have*

$$\boldsymbol{S}\big(\boldsymbol{\varphi} + \pi\,\boldsymbol{q}\big) = \boldsymbol{S}\big(\boldsymbol{\zeta}(\boldsymbol{q})\boldsymbol{\varphi}\big)\,\boldsymbol{D}(\boldsymbol{q}), \qquad \forall \boldsymbol{q} \in \mathbb{Z}^{k-1}. \tag{B.10}$$

That is, the effect of adding $\pi\,\boldsymbol{q}$ *to the angles can be expressed as (i) a componentwise sign flip of the angles determined by* $\boldsymbol{\zeta}(\boldsymbol{q})$*, and (ii) a right multiplication by a diagonal sign matrix* $\boldsymbol{D}(\boldsymbol{q})$*.*

Proof of Lemma B.5. Recall that $\boldsymbol{S}(\boldsymbol{\varphi}) = \prod_{i=2}^{k} \boldsymbol{R}_{i,1}(\varphi_i)$. We use two elementary identities.

(I1) For each $i \geqslant 2$, $\boldsymbol{R}_{i,1}(\varphi_i \pm \pi) = \boldsymbol{R}_{i,1}(\varphi_i)\boldsymbol{D}_{1i}$. This follows directly from the 2×2 rotation block in $\boldsymbol{R}_{i,1}(\cdot)$, since adding π multiplies that block by $-\boldsymbol{I}_2$.

(I2) For any $\boldsymbol{M} = \mathrm{diag}(m_1, \ldots, m_k)$, where $m_j = \pm 1$ for $j = 1, \ldots, k$, we have for any $i \geqslant 2$ that $\boldsymbol{M}\,\boldsymbol{R}_{i,1}(\varphi_i) = \boldsymbol{R}_{i,1}(m_1\, m_i\, \varphi_i)\,\boldsymbol{M}$.

Using (I1), we obtain $\boldsymbol{S}(\boldsymbol{\varphi} + \pi\,\boldsymbol{q}) = \prod_{i=2}^{k} \big(\boldsymbol{R}_{i,1}(\varphi_i)\boldsymbol{D}_{1i}^{q_i}\big)$. We now move all diagonal matrices $\boldsymbol{D}_{1i}^{q_i}$ to the far right. Each time a factor $\boldsymbol{D}_{1j}^{q_j}$ passes a later rotation $\boldsymbol{R}_{i,1}(\varphi_i)$ with $i > j$, identity (I2) shows that φ_i changes sign, because $\boldsymbol{D}_{1j}^{q_j}$ flips the first coordinate, but not the ith. Consequently, the total sign accumulated by φ_i equals $(-1)^{\sum_{j=2}^{i-1} q_j}$. That is, $\Big(\prod_{j=2}^{i-1} \boldsymbol{D}_{1j}^{q_j}\Big)\boldsymbol{R}_{i,1}(\varphi_i) = \boldsymbol{R}_{i,1}\Big((-1)^{\sum_{j=2}^{i-1} q_j}\,\varphi_i\Big)\Big(\prod_{j=2}^{i-1} \boldsymbol{D}_{1j}^{q_j}\Big)$. After all diagonal matrices have been moved to the right, they combine into $\boldsymbol{D}(\boldsymbol{q}) = \prod_{i=2}^{k} \boldsymbol{D}_{1i}^{q_i}$, and the remaining rotation product equals $\boldsymbol{S}\big(\boldsymbol{\zeta}(\boldsymbol{q})\boldsymbol{\varphi}\big)$. This proves (B.10). □

Lemma B.6. *Assume that for some vectors* $\boldsymbol{\varphi} = (\varphi_2, \ldots, \varphi_k)^\top$ *and* $\boldsymbol{\varphi}_0 = (\varphi_{0,2}, \ldots, \varphi_{0,k})^\top$*, we have* $\boldsymbol{S}(\boldsymbol{\varphi}) = \boldsymbol{S}(\boldsymbol{\varphi}_0)\,\boldsymbol{D}$*, where* $\boldsymbol{D} = \mathrm{diag}(\pm 1, \ldots, \pm 1)$ *is a sign matrix. Then there exists an integer vector* $\boldsymbol{q} = (q_2, \ldots, q_k)^\top \in \mathbb{Z}^{k-1}$ *such that*

$$\boldsymbol{\varphi} = \boldsymbol{\zeta}(\boldsymbol{q})\,\boldsymbol{\varphi}_0 + \pi\,\boldsymbol{q}, \tag{B.11}$$

that is, the angles satisfy the componentwise relation $\varphi_i = \zeta_i(\boldsymbol{q})\,\varphi_{0,i} + \pi\,q_i$ *for* $i = 2, \ldots, k$*, where* $\boldsymbol{\zeta}(\cdot) = \mathrm{diag}\big(\zeta_2(\cdot), \ldots, \zeta_k(\cdot)\big)$ *is given in Lemma B.5. Moreover,* $\boldsymbol{D} = \boldsymbol{D}(\boldsymbol{q}) = \prod_{i=2}^{k} \boldsymbol{D}_{1i}^{q_i}$*, where* $\boldsymbol{D}_{1i}$ *is defined in Lemma B.5.*

Proof of Lemma B.6. Note that $\det(\boldsymbol{D}) = 1$ must hold. Indeed, each Givens rotation matrix has determinant one, and hence $\det\big(\boldsymbol{S}(\boldsymbol{\varphi})\big) = \det\big(\boldsymbol{S}(\boldsymbol{\varphi}_0)\big) = 1$. Taking determinants on both sides of $\boldsymbol{S}(\boldsymbol{\varphi}) = \boldsymbol{S}(\boldsymbol{\varphi}_0)\,\boldsymbol{D}$ therefore gives $\det(\boldsymbol{D}) = 1$. We prove the representation (B.11) by induction on $k \geqslant 2$.

Consider $k = 2$. Then $\boldsymbol{S}(\boldsymbol{\varphi}) = \boldsymbol{R}_{2,1}(\varphi_2)$. We have $\boldsymbol{R}_{2,1}(\varphi_2) = \boldsymbol{R}_{2,1}(\varphi_{0,2})\,\boldsymbol{D}$. Since $\det(\boldsymbol{D}) = 1$, then $\boldsymbol{D} \in \{\boldsymbol{I}_2, -\boldsymbol{I}_2\}$. It is straightforward to verify that (B.11) holds with $\zeta_2(q_2) = 1$, with $q_2 \in \mathbb{Z}$, and that Lemma B.6 holds with $\boldsymbol{D} = (-\boldsymbol{I}_2)^{q_2}$.

Induction step. Assume that the statement holds for dimension $k-1$, and prove it for dimension $k \geqslant 3$. Write

$$\boldsymbol{S}(\boldsymbol{\varphi}) = \boldsymbol{R}_{2,1}(\varphi_2)\left(\prod_{i=3}^{k} \boldsymbol{R}_{i,1}(\varphi_i)\right).$$

A crucial fact is that for every $i \geqslant 3$, $\boldsymbol{R}_{i,1}(\cdot)$ does not mix coordinate 2 with any other coordinate; hence the second column of $\boldsymbol{S}(\boldsymbol{\varphi})$ depends only on φ_2 and equals $\mathrm{col}_2\left(\boldsymbol{S}(\boldsymbol{\varphi})\right) = (\sin(\varphi_2), \cos(\varphi_2), 0, \ldots, 0)^\top$, where $\mathrm{col}_j(\boldsymbol{Q})$ denotes the jth column of the matrix $\boldsymbol{Q}$. The same holds with $\boldsymbol{\varphi}_0$. Since right-multiplication by $\boldsymbol{D} = \mathrm{diag}(d_1, \ldots, d_k)$ multiplies the second column by d_2, it implies that $\sin(\varphi_2) = d_2 \sin(\varphi_{0,2})$ and $\cos(\varphi_2) = d_2 \cos(\varphi_{0,2})$. Therefore there exists $q_2 \in \mathbb{Z}$ such that

$$\varphi_2 = \varphi_{0,2} + \pi\, q_2, \qquad d_2 = (-1)^{q_2}. \tag{B.12}$$

Now left-multiply $\boldsymbol{S}(\boldsymbol{\varphi}) = \boldsymbol{S}(\boldsymbol{\varphi}_0)\,\boldsymbol{D}$ by $\boldsymbol{R}_{2,1}(-\varphi_2)$. By the basic angle sum and difference identities for trigonometric functions, one has

$$\boldsymbol{R}_{2,1}(-\varphi_2)\boldsymbol{R}_{2,1}(\varphi_2) = \boldsymbol{I}_k, \qquad \boldsymbol{R}_{2,1}(-\varphi_2)\boldsymbol{R}_{2,1}(\varphi_{0,2}) = \boldsymbol{R}_{2,1}(\varphi_{0,2}-\varphi_2) = \boldsymbol{R}_{2,1}(-\pi q_2).$$

Hence,

$$\prod_{i=3}^{k} \boldsymbol{R}_{i,1}(\varphi_i) = \boldsymbol{R}_{2,1}(-\pi q_2)\left(\prod_{i=3}^{k} \boldsymbol{R}_{i,1}(\varphi_{0,i})\right)\boldsymbol{D}. \tag{B.13}$$

By Part (I1) in the proof of Lemma B.5, we have $\boldsymbol{R}_{2,1}(-\pi q_2) = \boldsymbol{D}_{12}^{q_2}$. Define $\widetilde{\boldsymbol{D}} = \mathrm{diag}(\tilde{d}_1, \ldots, \tilde{d}_k) := \boldsymbol{D}_{12}^{q_2}\,\boldsymbol{D}$. Then $\widetilde{\boldsymbol{D}}$ is diagonal with $\det(\widetilde{\boldsymbol{D}}) = 1$, and its second diagonal entry is $\tilde{d}_2 = (-1)^{q_2} d_2 = 1$ by (B.12). By (B.13) and Part (I2) in the proof of Lemma B.5, we then obtain

$$\prod_{i=3}^{k} \boldsymbol{R}_{i,1}(\varphi_i) = \left(\prod_{i=3}^{k} \boldsymbol{R}_{i,1}\big(\tilde{\varphi}_{0,i}\big) \right) \widetilde{\boldsymbol{D}}, \qquad \tilde{\varphi}_{0,i} = (-1)^{q_2}\, \varphi_{0,i}. \tag{B.14}$$

Because the left-hand side of (B.14) leaves coordinate 2 unchanged, we may restrict (B.14) to the $(k-1)$-dimensional subspace spanned by coordinates $\{1, 3, 4, \ldots, k\}$ (i.e., remove the second row and second column). On that subspace, $\prod_{i=3}^{k} \boldsymbol{R}_{i,1}(\varphi_i)$ and $\prod_{i=3}^{k} \boldsymbol{R}_{i,1}(\varphi_{0,i})$ retain exactly the same structure as in dimension $k-1$ (with the role of index 3 now playing the role of "2", etc.), and the restricted $\widetilde{\boldsymbol{D}}^{(-2)}$ (i.e., obtained by removing the second row and second column of the original $\widetilde{\boldsymbol{D}}$) is again a diagonal sign matrix with determinant 1. Since (B.11) is assumed to hold for $k-1$ by the induction hypothesis, there exist integers $q_3, \ldots, q_k$ such that for each $i = 3, \ldots, k$,

$$\varphi_i = (-1)^{\sum_{j=3}^{i-1} q_j} \tilde{\varphi}_{0,i} + \pi\, q_i, \tag{B.15}$$

with the convention that $\sum_{j=\ell_2}^{\ell_1} \cdot \equiv 0$ for $\ell_1 < \ell_2$. Since $\tilde{\varphi}_{0,i} = (-1)^{q_2}\, \varphi_{0,i}$, we have $\varphi_i = (-1)^{\sum_{j=2}^{i-1} q_j}\, \varphi_{0,i} + \pi\, q_i$ for $i = 3, \ldots, k$, which, together with (B.12), implies that (B.11) also holds for k. Note that the additional factor $(-1)^{q_2}$ arises from the sign flip of the first coordinate introduced by $\widetilde{\boldsymbol{D}} = \boldsymbol{D}_{12}^{q_2} \boldsymbol{D}$ in (B.14). Hence, the induction step is established and therefore (B.11) holds.

Expression for $\boldsymbol{D}$. By (B.10) in Lemma B.5 and (B.11), we have $\boldsymbol{S}(\boldsymbol{\varphi}) = \boldsymbol{S}\big(\boldsymbol{\zeta}(\boldsymbol{q})\, \boldsymbol{\zeta}(\boldsymbol{q}) \boldsymbol{\varphi}_0\big) \boldsymbol{D}(\boldsymbol{q}) = \boldsymbol{S}(\boldsymbol{\varphi}_0) \boldsymbol{D}(\boldsymbol{q})$. On the other hand, $\boldsymbol{S}(\boldsymbol{\varphi}) = \boldsymbol{S}(\boldsymbol{\varphi}_0)\, \boldsymbol{D}$ by assumption. It then follows that $\boldsymbol{D} = \boldsymbol{D}(\boldsymbol{q})$, which completes the proof. □

C Additional empirical results

Figure C.1 complements Figure 6 in Section 5 by plotting the elements of the second eigenvector of the DSR model over time. Note that the second eigenvector automatically changes over time if we rotate the first eigenvector, at least if the first angle in $\boldsymbol{\varphi}_t$ is not zero. Differences, however, are less pronounced than for the first eigenvector; compare Figure 6. The loadings of the second eigenvector are positive for the top 5 institutions (less systemically important in terms of SRISK) and negative for the bottom 5 institutions (characterized by the highest average SRISK).

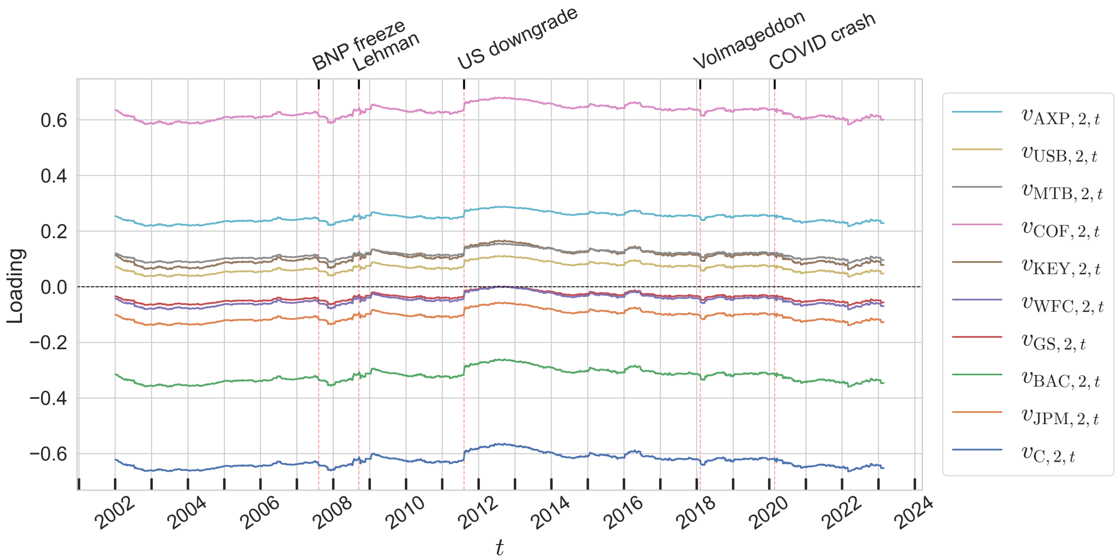

Figure C.1: Dynamics of the second eigenvector

This figure shows the time series of all elements of the second eigenvector of the conditional covariance matrix. The sample period spans January 2002 to January 2023.

As mentioned in Section 5, we also experimented with two alternative rotation schemes for the baseline orthogonal matrix $\bar{\boldsymbol{V}}$. The first specification, labeled DSR Δ_3, only allows for rotations between the eigenvectors corresponding to the three most dominant eigenvalues. Namely, we set

$$\boldsymbol{V}_t = \bar{\boldsymbol{V}} \prod_{j=1}^{2} \prod_{i=j+1}^{3} \boldsymbol{R}_{i,j}(\varphi_{i,j,t}), \tag{C.1}$$

Table C.1: Additional DSR specifications

This table shows the results for 10 financial US equities between various DSR specifications. In particular, the table includes the DSR model adopted in Section 5, a specification labeled "DSR $\triangle_3$" defined in (C.1), and a specification labeled "DSR 2vec" in (C.2). The results are based on predictions on 16 years of data spanning from January 2007 to February 2023. The columns represent the performance measures: the in-sample log-likelihood, the out-of-sample log-score, the mean Frobenius-normed difference between the forecasts of the conditional covariance matrix and the observed (estimated) realized covariance matrix, and the QLIKE value. Stars indicate statistical significance of the DM test against the DSR model for the out-of-sample measures: *$p < 0.10$, **$p < 0.05$, and ***$p < 0.01$.

	In-sample Log-Lik	Log-score	Frobenius Norm	QLIKE
DSR	−74 324	−14.588	23.735	14.323
DSR $\triangle_3$	−74 966	−14.679***	24.003***	14.274
DSR 2vec	−74 048	−14.682***	23.967***	14.497***

where $\varphi_{i,j,t}$ is the rotation angle in the (i,j) plane. The idea behind this specification is that most of the empirical gain can be obtained by rotations along the first three principal axes, as these capture most of the variation in the data. In our second alternative specification, labeled DSR 2vec, we rotate both the first and the second eigenvectors (with all other possible eigenvectors), i.e.,

$$\boldsymbol{V}_t = \bar{\boldsymbol{V}} \prod_{j=1}^{2} \prod_{i=j+1}^{k} \boldsymbol{R}_{i,j}(\varphi_{i,j,t}). \tag{C.2}$$

The idea behind this specification is that changing the second principal axis over time may also result in sizable empirical gains, while allowing the rotations to have maximum freedom along all principal axes.

Table C.1 shows that the baseline DSR specification in Section 5 captures the data better than the two alternative specifications. Although in-sample the DSR 2vec model performs better than the baseline model, this appears to be due to overfitting. When we move to the out-of-sample results, the baseline model appears to be the overall winner. Only for QLIKE, the DSR Δ_3 specification does slightly (insignificantly) better, whereas the model loses in terms of log-score and Frobenius norm. Given these results, we restrict ourselves to the baseline DSR specification.

References in Online Appendix

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.

Berkes, I., L. Horváth, and P. Kokoszka (2003). GARCH processes: Structure and estimation. *Bernoulli*, 201–227.

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.

Davidson, J. (1994). *Stochastic Limit Theory.* Oxford: Oxford University Press.

Kechris, A. (2012). *Classical Descriptive Set Theory*, Volume 156. New York: Springer.

Lin, Y. and A. Lucas (2025a). Functional location-scale models with robust observation-driven dynamics. Technical Report TI 2025-027/III, Tinbergen Institute Discussion Paper.

Lin, Y. and A. Lucas (2025b). Testing for the absence of score-driven parameter dynamics. Technical Report TI 2025-063/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.

Magnus, J. R. and H. Neudecker (2019). *Matrix Differential Calculus with Applications in Statistics and Econometrics.* John Wiley & Sons.

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: Cambridge University Press.

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