Testing for observation-dependent regime switching in mixture autoregressive models*

Mika Meitz
University of Helsinki

Pentti Saikkonen
University of Helsinki

October 2017

Abstract

Testing for regime switching when the regime switching probabilities are specified either as constants ('mixture models') or are governed by a finite-state Markov chain ('Markov switching models') are long-standing problems that have also attracted recent interest. This paper considers testing for regime switching when the regime switching probabilities are time-varying and depend on observed data ('observation-dependent regime switching'). Specifically, we consider the likelihood ratio test for observation-dependent regime switching in mixture autoregressive models. The testing problem is highly nonstandard, involving unidentified nuisance parameters under the null, parameters on the boundary, singular information matrices, and higher-order approximations of the log-likelihood. We derive the asymptotic null distribution of the likelihood ratio test statistic in a general mixture autoregressive setting using high-level conditions that allow for various forms of dependence of the regime switching probabilities on past observations, and we illustrate the theory using two particular mixture autoregressive models. The likelihood ratio test has a nonstandard asymptotic distribution that can easily be simulated, and Monte Carlo studies show the test to have satisfactory finite sample size and power properties.

JEL classification: C12, C22, C52.

Keywords: Likelihood ratio test, singular information matrix, higher-order approximation of the log-likelihood, logistic mixture autoregressive model, Gaussian mixture autoregressive model.

*The authors thank the Academy of Finland for financial support. Contact addresses: Mika Meitz, Department of Political and Economic Studies, University of Helsinki, P. O. Box 17, FI–00014 University of Helsinki, Finland; e-mail: mika.meitz@helsinki.fi. Pentti Saikkonen, Department of Mathematics and Statistics, University of Helsinki, P. O. Box 68, FI–00014 University of Helsinki, Finland; e-mail: pentti.saikkonen@helsinki.fi.

1 Introduction

Different regime switching models are in widespread use in economics, finance, and other fields. When the regime switching probabilities are constants, these models are often referred to as 'mixture models', and when these probabilities depend on past regimes and are governed by a finite-state Markov chain, the term (time homogeneous) 'Markov switching models' is typically used. In this paper, we are interested in the case where the regime switching probabilities depend on observed data, a case we refer to as 'observation-dependent regime switching'. Models of this kind can be viewed as special cases of time inhomogeneous Markov switching models (in which regime switching probabilities depend on both past regimes and observed data). Overviews of regime switching models can be found, for example, in Frühwirth-Schnatter (2006) and Hamilton (2016). Of critical interest in all these models is whether the use of several regimes is warranted or if a single-regime model would suffice. Testing for regime switching in all these models is plagued by several irregular features such as unidentified parameters and parameters on the boundary and is consequently notoriously difficult.

Tests for Markov switching have been considered by several authors in the econometrics literature. Hansen (1992) and Garcia (1998) both considered sup-type likelihood ratio (LR) tests in Markov switching models but they did not present complete solutions. Hansen derived a bound for the distribution of the LR statistic, leading to a conservative procedure, while Garcia did not treat all the non-standard features of the problem in detail. Cho and White (2007) analyzed the use of a LR statistic for a mixture model to test for Markov-switching type regime switching. They found their test based on a mixture model to have power against Markov switching alternatives even though it ignores the temporal dependence of the Markov chain. Carrasco, Hu, and Ploberger (2014) took a different approach and proposed an information matrix type test that they showed to be asymptotically optimal against Markov switching alternatives. Very recently, both Qu and Zhuo (2017) and Kasahara and Shimotsu (2017) have studied the LR statistic for regime switching in Markov switching models.

Regarding testing for mixture type regime switching, the existing literature is extensive, and several early references can be found, for instance, in McLachlan and Peel (2000, Sec. 6.5.1). Most papers in this literature consider the case of independent observations without regressors. Notable exceptions allowing for regressors (but not dependent data) and having set-ups closer to the present paper are Zhu and Zhang (2004, 2006) and Kasahara and Shimotsu (2015) who consider (among other things) LR tests for regime switching. Further comparison to these works will be provided in later sections.

In contrast to testing for Markov switching or mixture type regime switching, there exists almost no literature on testing for observation-dependent regime switching. The only two exceptions we are aware of are the unpublished PhD thesis of Jeffries (1998) and the recent paper of Shen and He (2015). Jeffries's thesis, which appears to have gone largely unnoticed, analyses the LR test in a specific (first-order) mixture autoregressive model; we will discuss his work further in later sections. Shen and He (2015) consider the case of independent observations with regressors and observation-dependent regime switching, and propose an 'expectation maximization test' for regime switching.

In this paper we consider testing for observation-dependent regime switching in a time series context. Specifically, we analyze the asymptotic distribution of the LR test statistic for testing a linear autoregressive model against a two-regime mixture autoregressive model with observation-dependent regime switching. Mixture autoregressive models have been discussed for instance in Wong and Li (2000, 2001), Dueker, Sola, and Spagnolo (2007), Dueker, Psaradakis, Sola, and Spagnolo (2011), and Kalliovirta, Meitz, and Saikkonen (2015, 2016); further discussion of this previous work will be provided in Section 2. Motivation for allowing the regime switching probabilities to depend on observed data stems, for instance, from the desire to associate changes in regime to observable economic variables.[1,2] Following Kasahara and Shimotsu (2015) it would also be possible to consider the more general testing problem that in a model with more than two regimes the number of regimes can be reduced. However, as even the case of two regimes is quite complex in our set-up, we leave this extension to future research.

We consider mixture autoregressive (MAR) models in a rather general setting employing high-level conditions that allow for various forms of observation-dependent regime switching. As specific examples, we treat the so-called logistic MAR (LMAR) model of Wong and Li (2001) and (a version of the) Gaussian MAR (GMAR) model of Kalliovirta et al. (2015) in detail. The technical challenges we face in analyzing the LR test statistic are similar to those when testing for Markov switching and mixture type regime switching. First, there are nuisance parameters that are unidentified under the null hypothesis. This is the classical Davies (1977, 1987) type problem. Second, under the null hypothesis, there are parameters on the boundary of the permissible parameter space. Such problems (also allowing for unidentified nuisance parameters under the null) are discussed in Andrews (1999, 2001). Third, the Fisher information matrix is (potentially) singular, preventing the use of conventional second-order expansions of the log-likelihood to analyze the LR test statistic. Such problems are discussed by Rotnitzky, Cox, Bottai, and Robins (2000), and suitable reparameterizations and higher-order expansions are needed to analyze the LR statistic. A particular challenge in the present paper is to deal with these three problems simultaneously. Similar problems were faced by Kasahara and Shimotsu (2015), and inspired by their work we consider a suitably reparameterized model, write a higher-order expansion of the log-likelihood function as a quadratic function of the new parameters, and then derive the asymptotic distribution of the LR test statistic by slightly extending and adapting the arguments of Andrews (1999, 2001) and Zhu and Zhang (2006) (who partially generalize results of Andrews). Our two examples demonstrate that, compared to the mixture type regime switching considered by Kasahara and Shimotsu (2015), observation-dependent regime switching can either simplify or complicate the analysis of the LR test statistic.

We contribute to the literature in several ways. (1) To the best of our knowledge, apart from the unpublished PhD thesis of Jeffries (1998), we are the first to study testing for observation-dependent regime switching using the LR test statistic and among the rather few to allow for dependent observations. (2) We provide a general framework to cover various forms of observation-dependent regime switching, making our results potentially applicable to several models not explicitly discussed in the present paper. (3) From a methodological perspective, we slightly extend and adapt certain arguments of Andrews (1999, 2001) and Zhu and Zhang (2006), which may be of independent interest.

The rest of the paper is organized as follows. Section 2 reviews mixture autoregressive models. Section 3 analyzes the LR test statistic for testing a linear autoregressive model against a two-regime mixture autoregressive model. Simulation-based critical values and a Monte Carlo study are discussed in Section 4, and Section 5 concludes. Appendices A–C contain technical details and proofs. Supplementary Appendices D–E, available from the authors upon request, contain further technical details omitted from the paper.

Finally, a few notational conventions are given. All vectors will be treated as column vectors and, for the sake of uncluttered notation, we shall write $x = (x_1, \ldots, x_n)$ for the (column) vector x where the

[1] See, for instance, Hamilton (2016), whose *Handbook of Macroeconomics* chapter begins "Many economic time series exhibit dramatic breaks associated with events such as economic recessions, financial panics, and currency crises. Such changes in regime may arise from tipping points or other nonlinear dynamics and are core to some of the most important questions in macroeconomics."

[2] More general models in which the regime switching probabilities are allowed to depend on both past regimes and observable variables have also been considered, see, e.g., Diebold, Lee, and Weinbach (1994), Filardo (1994), and Kim, Piger, and Startz (2008).

components x_i may be either scalars or vectors (or both). For any vector or matrix x, the Euclidean norm is denoted by $\|x\|$. We let $X_{T\alpha} = o_{p\alpha}(1)$ and $X_{T\alpha} = O_{p\alpha}(1)$ stand for $\sup_{\alpha \in A} \|X_{T\alpha}\| = o_p(1)$ and $\sup_{\alpha \in A} \|X_{T\alpha}\| = O_p(1)$, respectively, and $\lambda_{\min}(\cdot)$ and $\lambda_{\max}(\cdot)$ for the smallest and largest eigenvalue of the indicated matrix.

2 Mixture autoregressive models

2.1 General formulation

Let y_t $(t = 1, 2, \ldots)$ be a real-valued time series of interest, and let $\mathcal{F}_{t-1} = \sigma(y_s,\ s < t)$ denote the σ–algebra generated by past y_t's. We use $P_{t-1}(\cdot)$ to signify the conditional probability of the indicated event given $\mathcal{F}_{t-1}$. In the general two component mixture autoregressive model we consider the y_t's are generated by

$$y_t = s_t\left(\tilde{\phi}_0 + \sum_{i=1}^{p} \tilde{\phi}_i y_{t-i} + \tilde{\sigma}_1 \varepsilon_t\right) + (1 - s_t)\left(\tilde{\varphi}_0 + \sum_{i=1}^{p} \tilde{\varphi}_i y_{t-i} + \tilde{\sigma}_2 \varepsilon_t\right), \tag{1}$$

where the parameters $\tilde{\sigma}_1$ and $\tilde{\sigma}_2$ are positive, and conditions required for the autoregressive parameters $\tilde{\phi}_i$ and $\tilde{\varphi}_i$ $(i = 1, \ldots, p)$ will be discussed later. Furthermore, ε_t and s_t are (unobserved) stochastic processes which satisfy the following conditions: (a) ε_t is a sequence of independent standard normal random variables such that ε_t is independent of $\{y_{t-j},\ j > 0\}$, (b) s_t is a sequence of Bernoulli random variables such that, for each t, $P_{t-1}(s_t = 1) = \alpha_t$ with α_t a function of $\boldsymbol{y}_{t-1} = (y_{t-1}, \ldots, y_{t-p})$, and (c) conditional on $\mathcal{F}_{t-1}$, s_t and ε_t are independent.

The conditional probabilities α_t and $1-\alpha_t$ $(= P_{t-1}(s_t = 0))$ are referred to as mixing weights. They can be thought of as (conditional) probabilities that determine which one of the two autoregressive components of the mixture generates the next observation y_t. In condition (b) it is assumed that the mixing weight α_t (and hence also the conditional distribution of y_t given its past) only depends on p lags of y_t; allowing for more than p lags in the mixing weight would be possible at the cost of more complicated notation.

We assume that of the original parameters $\tilde{\phi} = (\tilde{\phi}_0, \tilde{\phi}_1, \ldots, \tilde{\phi}_p, \tilde{\sigma}_1^2)$ and $\tilde{\varphi} = (\tilde{\varphi}_0, \tilde{\varphi}_1, \ldots, \tilde{\varphi}_p, \tilde{\sigma}_2^2)$ in the two regimes, q_1 parameters are a priori assumed the same in both regimes and the remaining q_2 parameters are potentially different in the two regimes (with $q_1 + q_2 = p + 2$). For instance, one may assume that $\tilde{\phi}_0$ and $\tilde{\varphi}_0$ are equal, or alternatively that $\tilde{\sigma}_1^2$ and $\tilde{\sigma}_2^2$ are equal. If such an assumption is plausible, taking it into account when devising a test for regime switching will be advantageous (it will lead to a test with better power). To this end, let β be a $q_1 \times 1$ vector of common parameters, and let ϕ and φ be $q_2 \times 1$ vectors of (potentially) different parameters. Then, for some known $(p+2)$–dimensional permutation matrix P, $(\beta, \phi) = P\tilde{\phi}$ and $(\beta, \varphi) = P\tilde{\varphi}$. For simplicity, we assume that β and ϕ are variation-free, requiring the autoregressive coefficients $\tilde{\phi}_1, \ldots, \tilde{\phi}_p$ to be contained in either β or ϕ (the same variation-freeness is assumed of β and φ). If there are no common coefficients in the two regimes, the parameter β can be dropped and $\phi = \tilde{\phi}$ and $\varphi = \tilde{\varphi}$.

As for the mixing weight α_t, in addition to past y_t's it depends on unknown parameters which may include components of the parameter vector (β, ϕ, φ) and an additional parameter α (scalar or vector). When this dependence needs to be emphasized we use the notation $\alpha_t(\alpha, \beta, \phi, \varphi)$.

Using equation (1) and the conditions following it, the conditional density function of y_t given its past, $f(\cdot \mid \mathcal{F}_{t-1})$, is obtained as

$$f(y_t \mid \mathcal{F}_{t-1}) = \alpha_t f_t(\beta, \phi) + (1 - \alpha_t) f_t(\beta, \varphi), \tag{2}$$

where the notation $f_t(\beta, \phi)$ signifies the density function of a (univariate) normal distribution with mean $\tilde{\phi}_0 + \tilde{\phi}_1 y_{t-1} + \cdots + \tilde{\phi}_p y_{t-p}$ and variance $\tilde{\sigma}_1^2$ evaluated at y_t, that is,

$$f_t(\beta, \phi) = \frac{1}{\tilde{\sigma}_1} \mathfrak{n}\left(\frac{y_t - (\tilde{\phi}_0 + \tilde{\phi}_1 y_{t-1} + \cdots + \tilde{\phi}_p y_{t-p})}{\tilde{\sigma}_1} \right), \tag{3}$$

with $\mathfrak{n}(u) = (2\bar{\pi})^{-1/2} \exp(-u^2/2)$ the density function of a standard normal random variable and $\bar{\pi} = 3.14\ldots$ the number pi. The notation $f_t(\beta, \varphi)$ is defined similarly by using the parameters $\tilde{\varphi}_i$ ($i = 0, \ldots, p$) and $\tilde{\sigma}_2^2$ instead of $\tilde{\phi}_i$ ($i = 0, \ldots, p$) and $\tilde{\sigma}_1^2$. Thus, the distribution of y_t given its past is specified as a mixture of two normal densities with time varying mixing weights α_t and $1 - \alpha_t$.

Different mixture autoregressive models are obtained by different specifications of the mixing weights (or in our case the single mixing weight α_t). In some of the proposed models more than two mixture components are allowed but for reasons to be discussed below we will not consider these extensions. If the mixing weights are assumed constant over time the general mixture autoregressive model introduced above reduces to (a two component version) of the MAR model studied by Wong and Li (2000). In the LMAR model of Wong and Li (2001), a logistic transformation of the two mixing weights is assumed to be a linear function of past observed variables. Related two-regime mixture models with time-varying mixing weights have also been considered by Gouriéroux and Robert (2006), Dueker et al. (2007) and Bec, Rahbek, and Shephard (2008) whereas Lanne and Saikkonen (2003) and Kalliovirta et al. (2015) have considered mixture autoregressive models in which multiple regimes are allowed.

A common problem with the application of mixture autoregressive models is determining the value of the (usually unknown) number of component models or regimes. As discussed in the Introduction, several irregular features make this problem difficult and these difficulties are encountered even when the observations are a random sample from a mixture of (two) normal distributions. To our knowledge the only solution presented for mixture autoregressive models is provided for a simple first order case with no intercept terms in the unpublished PhD thesis of Jeffries (1998). As discussed in the recent papers by Kasahara and Shimotsu (2012, 2015) and the references therein, some of the difficulties involved stem from properties of the normal distribution.

The difficulties referred to above also partly explain the complexity of our testing problem and why we only consider test procedures that can be used to test the null hypothesis that a two component mixture autoregressive model reduces to a conventional linear autoregressive model. Following the ideas in Zhu and Zhang (2006) and Kasahara and Shimotsu (2012, 2015), we derive a LR test in the general set-up described above and apply it to two particular cases, the LMAR model of Wong and Li (2001) and the GMAR model of Kalliovirta et al. (2015). Next, we shall discuss these two models in more detail.

2.2 Two particular examples

LMAR Example. The LMAR model of Wong and Li (2001) is defined by specifying the mixing weight α_t as

$$\alpha_t^L = \alpha_t^L(\alpha) = \frac{\exp(\alpha_0 + \alpha_1 y_{t-1} + \cdots + \alpha_r y_{t-m})}{1 + \exp(\alpha_0 + \alpha_1 y_{t-1} + \cdots + \alpha_r y_{t-m})},$$

where the vector $\alpha = (\alpha_0, \alpha_1, \ldots, \alpha_m)$ contains $m+1$ unknown parameters and the order m ($1 \leq m \leq p$) is assumed known.

GMAR Example. In the GMAR model of Kalliovirta et al. (2015) the mixing weight is defined as

$$\alpha_t^G = \alpha_t^G(\alpha, \tilde{\phi}, \tilde{\varphi}) = \frac{\alpha \mathsf{n}_p(\boldsymbol{y}_{t-1}; \tilde{\phi})}{\alpha \mathsf{n}_p(\boldsymbol{y}_{t-1}; \tilde{\phi}) + (1-\alpha)\mathsf{n}_p(\boldsymbol{y}_{t-1}; \tilde{\varphi})}, \tag{4}$$

where $\alpha \in (0,1)$ is an unknown parameter and $\mathsf{n}_p(\boldsymbol{y}_{t-1}; \cdot)$ denotes the density function of a particular p–dimensional ($p \geq 1$) normal distribution defined as follows.

First, define the auxiliary Gaussian AR(p) processes (cf. equation (1))

$$\nu_{1,t} = \tilde{\phi}_0 + \sum_{i=1}^{p} \tilde{\phi}_i \nu_{1,t-i} + \tilde{\sigma}_1 \varepsilon_t \quad \text{and} \quad \nu_{2,t} = \tilde{\varphi}_0 + \sum_{i=1}^{p} \tilde{\varphi}_i \nu_{2,t-i} + \tilde{\sigma}_2 \varepsilon_t,$$

where the autoregressive coefficients are assumed to satisfy

$$\tilde{\phi}(z) := 1 - \sum_{i=1}^{p} \tilde{\phi}_i z^i \neq 0 \text{ for } |z| \leq 1 \text{ and } \tilde{\varphi}(z) := 1 - \sum_{i=1}^{p} \tilde{\varphi}_i z^i \neq 0 \text{ for } |z| \leq 1. \tag{5}$$

This condition implies that the processes $\nu_{1,t}$ and $\nu_{2,t}$ are stationary and that each of the two component models in (1) satisfies the usual stationarity condition of the conventional linear AR(p) model. Now set $\boldsymbol{\nu}_{m,t} = (\nu_{m,t}, \ldots, \nu_{m,t-p+1})$ and $\mathbf{1}_p = (1, \ldots, 1)$ ($p \times 1$), and let $\mu_m \mathbf{1}_p$ and $\boldsymbol{\Gamma}_{m,p}$ signify the mean vector and covariance matrix of $\boldsymbol{\nu}_{m,t}$ ($m = 1, 2$).[3] The random vector $\boldsymbol{\nu}_{1,t}$ follows the p–dimensional multivariate normal distribution with density

$$\mathsf{n}_p(\boldsymbol{\nu}_{1,t}; \tilde{\phi}) = (2\bar{\pi})^{-p/2} \det(\boldsymbol{\Gamma}_{1,p})^{-1/2} \exp\left\{ -\tfrac{1}{2} (\boldsymbol{\nu}_{1,t} - \mu_1 \mathbf{1}_p)' \boldsymbol{\Gamma}_{1,p}^{-1} (\boldsymbol{\nu}_{1,t} - \mu_1 \mathbf{1}_p) \right\}, \tag{6}$$

and the density of $\boldsymbol{\nu}_{2,t}$, denoted by $\mathsf{n}_p(\boldsymbol{\nu}_{2,t}; \tilde{\varphi})$, is defined similarly. Equation (1) and conditions (4)–(6) define the (two component) GMAR model (condition (5) is part of the definition of the model because it is used to define the mixing weights).

3 Test procedure

We now consider a test procedure of the null hypothesis that a two component mixture autoregressive model reduces to a conventional linear autoregressive model.

3.1 The null hypothesis and the LR test statistic

We denote the conditional density function corresponding to the unrestricted model as (see (2))

$$f_{2,t}(\alpha, \beta, \phi, \varphi) := f_2(y_t \mid \boldsymbol{y}_{t-1}; \alpha, \beta, \phi, \varphi) := \alpha_t(\alpha, \beta, \phi, \varphi) f_t(\beta, \phi) + (1 - \alpha_t(\alpha, \beta, \phi, \varphi)) \, f_t(\beta, \varphi),$$

where we now make the dependence of the mixing weight on the parameters explicit. With this notation the log-likelihood function of the model based on a sample $(y_{-p+1}, \ldots, y_T)$ (and conditional on the initial values $(y_{-p+1}, \ldots, y_0)$) is $L_T(\alpha, \beta, \phi, \varphi) = \sum_{t=1}^{T} l_t(\alpha, \beta, \phi, \varphi)$ where

$$l_t(\alpha, \beta, \phi, \varphi) = \log[f_{2,t}(\alpha, \beta, \phi, \varphi)] = \log\left[\alpha_t(\alpha, \beta, \phi, \varphi) f_t(\beta, \phi) + (1 - \alpha_t(\alpha, \beta, \phi, \varphi)) \, f_t(\beta, \varphi)\right].$$

[3] We have $\mu_1 = \tilde{\phi}_0 / \tilde{\phi}(1)$ and $\mu_2 = \tilde{\varphi}_0 / \tilde{\varphi}(1)$, whereas each of $\boldsymbol{\Gamma}_{m,p}$, ($m = 1, 2$), has the familiar form of being a $p \times p$ symmetric Toeplitz matrix with $\gamma_{m,0} = Cov[\nu_{m,t}, \nu_{m,t}]$ along the main diagonal, and $\gamma_{m,i} = Cov[\nu_{m,t}, \nu_{m,t-i}]$, $i = 1, \ldots, p-1$, on the diagonals above and below the main diagonal. Similarly to μ_1 and μ_2 the elements of the covariance matrices $\boldsymbol{\Gamma}_{1,p}$ and $\boldsymbol{\Gamma}_{2,p}$ are treated as functions of the parameters $\tilde{\phi}$ and $\tilde{\varphi}$, respectively (for details of this dependence, see Lütkepohl (2005, eqn. (2.1.39))).

The following assumption provides conditions on the data generation process, the parameter space of $(\alpha, \beta, \phi, \varphi)$, and the mixing weight $\alpha_t(\alpha, \beta, \phi, \varphi)$.

Assumption 1.

(i) *The y_t's are generated by a stationary linear Gaussian AR(p) model with (the true but unknown) parameter value $\tilde{\phi}^*$ an interior point of $\tilde{\Phi}$, a compact subset of $\{\tilde{\phi} = (\tilde{\phi}_0, \tilde{\phi}_1, \ldots, \tilde{\phi}_p, \tilde{\sigma}^2) \in \mathbb{R}^{p+2} : \tilde{\phi}_0 \in \mathbb{R};\ 1 - \sum_{i=1}^{p} \tilde{\phi}_i z^i \neq 0 \ \ for\ |z| \leq 1;\ \tilde{\sigma}^2 \in (0, \infty)\}$.*

(ii) *The parameter space of $(\alpha, \beta, \phi, \varphi)$ is $A \times B \times \Phi \times \Phi$, where A is a compact subset of $\mathbb{R}^a$ and B and Φ are those compact subsets of $\mathbb{R}^{q_1}$ and $\mathbb{R}^{q_2}$, respectively, that satisfy $(\beta, \phi) \in B \times \Phi$ if and only if $P^{-1}(\beta, \phi) \in \tilde{\Phi}$ (here P is as in the third paragraph of Section 2.1).*

(iii) *For all t and all $(\alpha, \beta, \phi, \varphi) \in A \times B \times \Phi \times \Phi$, the mixing weight $\alpha_t(\alpha, \beta, \phi, \varphi)$, is $\sigma(\boldsymbol{y}_{t-1})$–measurable (with $\sigma(\boldsymbol{y}_{t-1})$ denoting the σ–algebra generated by $\boldsymbol{y}_{t-1}$) and satisfies $\alpha_t(\alpha, \beta, \phi, \varphi) \in (0, 1)$.*

As our interest is to study the asymptotic null distribution of the LR test statistic, Assumption 1(i) requires the data to be generated by a stationary linear Gaussian AR(p) model. Assuming a compact parameter space in Assumptions 1(i) and (ii) is a standard requirement which facilitates proofs. Assumption 1(ii) accommodates to the main cases of interest, namely $\beta = \tilde{\phi}_0$, $\beta = (\tilde{\phi}_1, \ldots, \tilde{\phi}_p)$, $\beta = \tilde{\sigma}^2$, or any combination of these.[4]

Assumption 1(iii) implies that our two-component mixture autoregressive model reduces to a linear autoregression only when $\phi = \varphi$, regardless of the values of $\alpha \in A$ and $\beta \in B$. The null hypothesis to be tested is therefore $\phi = \varphi$ and the alternative is $\phi \neq \varphi$ or, more precisely,

$$H_0 : (\phi, \varphi) \in (\Phi \times \Phi)^0,\ \alpha \in A,\ \beta \in B \qquad \text{vs.} \qquad H_1 : (\phi, \varphi) \in (\Phi \times \Phi) \setminus (\Phi \times \Phi)^0,\ \alpha \in A,\ \beta \in B,$$

where

$$(\Phi \times \Phi)^0 = \{(\phi, \varphi) \in \Phi \times \Phi : \phi = \varphi\}.$$

Note that under the null hypothesis the parameter α vanishes from the likelihood function and is therefore unidentified.

Let $f_t^0(\tilde{\phi}) := f^0(y_t \mid \boldsymbol{y}_{t-1}; \tilde{\phi})$ and $L_T^0(\tilde{\phi})$ denote the conditional density and log-likelihood corresponding to the restricted model, that is,

$$f_t^0(\tilde{\phi}) = f_{2,t}(\alpha, \beta, \phi, \phi) = f_t(\tilde{\phi}) \quad \text{and} \quad L_T^0(\tilde{\phi}) = \sum_{t=1}^{T} l_t^0(\tilde{\phi}) \quad \text{with} \quad l_t^0(\tilde{\phi}) = \log[f_t(\tilde{\phi})]$$

(here the superscript 0 refers to the model restricted by the null hypothesis). Note that these quantities are obtained from a linear Gaussian AR(p) model. As $f_{2,t}(\alpha, \beta^*, \phi^*, \phi^*) = f_t(\tilde{\phi}^*)$ for any $\alpha \in A$, in the unrestricted model the parameter vector $(\alpha, \beta^*, \phi^*, \phi^*)$ corresponds to the true model for any $\alpha \in A$.

As already indicated, Assumption 1(iii) implies that the restriction $\phi = \varphi$ is the only possibility to formulate the null hypothesis. However, this is not necessarily the case if (against Assumption 1(iii)) the mixing weight $\alpha_t(\alpha, \beta, \phi, \varphi)$ were allowed to take the boundary values zero and one. Of our two examples this would be possible for the GMAR model of Kalliovirta et al. (2015) but not for the LMAR model of Wong and Li (2001). In the GMAR model $\alpha_t(\alpha, \beta, \phi, \varphi)$ takes the boundary values zero and one when the parameter α takes these values (see Section 2.2). In both cases a linear autoregression results and either the parameter ϕ or φ is unidentified (see (2)) (the MAR model of Wong and Li

[4]Note that assuming the autoregressive parameters ϕ and φ to have a common parameter space is made for simplicity and could be relaxed; for an example where such a relaxation would be needed, see the ACR model of Bec et al. (2008).

(2000) provides a similar example). It would be possible to obtain tests for the GMAR model by using the null hypotheses which specifies $\alpha = 0$ or $\alpha = 1$. However, as in Kasahara and Shimotsu (2012) (see also Kasahara and Shimotsu (2015)), this approach would require rather restrictive assumptions and would also lead to very complicated derivations.[5] Therefore, we will not consider this option.

As the parameter α is unidentified under the null hypothesis, the appropriate likelihood ratio type test statistic is

$$LR_T = \sup_{\alpha \in A} LR_T(\alpha),$$

where, for each fixed $\alpha \in A$,

$$LR_T(\alpha) = 2[\sup_{(\beta,\phi,\varphi)\in B\times\Phi\times\Phi} L_T(\alpha,\beta,\phi,\varphi) - \sup_{\tilde{\phi}\in\tilde{\Phi}} L_T^0(\tilde{\phi})].$$

To obtain an operational test statistic let, for each fixed $\alpha \in A$, $(\hat{\beta}_{T\alpha}, \hat{\phi}_{T\alpha}, \hat{\varphi}_{T\alpha})$ denote an (approximate) unrestricted maximum likelihood (ML) estimator of the parameter vector (β, ϕ, φ). We make the following assumption.

Assumption 2. *The unrestricted ML estimator satisfies the following conditions:*

(i) $L_T(\alpha, \hat{\beta}_{T\alpha}, \hat{\phi}_{T\alpha}, \hat{\varphi}_{T\alpha}) = \sup_{(\beta,\phi,\varphi)\in B\times\Phi\times\Phi} L_T(\alpha,\beta,\phi,\varphi) + o_{p\alpha}(1)$,

(ii) $(\hat{\beta}_{T\alpha}, \hat{\phi}_{T\alpha}, \hat{\varphi}_{T\alpha}) = (\beta^*, \phi^*, \phi^*) + o_{p\alpha}(1)$.

Assumption 2(i) means that $(\hat{\beta}_{T\alpha}, \hat{\phi}_{T\alpha}, \hat{\varphi}_{T\alpha})$ is assumed to maximize the likelihood function only asymptotically. This assumption is technical and made for ease of exposition (see Andrews (1999) and Zhu and Zhang (2006) for similar assumptions in related problems). Assumption 2(ii) is a high level condition on (uniform) consistency of the ML estimator and is analogous to Assumption 1 of Andrews (2001). It has to be verified on a case by case basis (this is exemplified below for the LMAR model and GMAR model).

As for the term $\sup_{\tilde{\phi}\in\tilde{\Phi}} L_T^0(\tilde{\phi})$ in the LR test statistic, note that $L_T^0(\tilde{\phi})$ is the (conditional) log-likelihood function of a linear Gaussian AR(p) model. Let $\hat{\tilde{\phi}}_T$ denote an (approximate) maximum likelihood estimator of the parameters of a linear Gaussian AR(p) model, that is, $\hat{\tilde{\phi}}_T$ satisfies[6]

$$L_T^0(\hat{\tilde{\phi}}_T) = \sup_{\tilde{\phi}\in\tilde{\Phi}} L_T^0(\tilde{\phi}) + o_p(1) \text{ and } \hat{\tilde{\phi}}_T = \tilde{\phi}^* + o_p(1).$$

Noting that $L_T(\alpha, \beta^*, \phi^*, \phi^*) = L_T^0(\tilde{\phi}^*)$ for any α now allows us to write $LR_T(\alpha)$ as

$$LR_T(\alpha) = 2[L_T(\alpha, \hat{\beta}_{T\alpha}, \hat{\phi}_{T\alpha}, \hat{\varphi}_{T\alpha}) - L_T(\alpha, \beta^*, \phi^*, \phi^*)] - 2[L_T^0(\hat{\tilde{\phi}}_T) - L_T^0(\tilde{\phi}^*)] + o_{p\alpha}(1). \quad (7)$$

The analysis of the second term on the right hand side is standard while dealing with the first term is more demanding requiring a substantial amount of preparation.

3.1.1 Examples (continued)

LMAR Example. In the LMAR example, we assume there are no common parameters in the two regimes so that the parameter β is omitted, $\phi = \tilde{\phi}$, $\varphi = \tilde{\varphi}$, $q_1 = 0$, and $q_2 = p + 2$. To satisfy

[5] See, for instance, the remarks following Proposition 5 in Kasahara and Shimotsu (2012) or property (a) on p. 1633 of Kasahara and Shimotsu (2015).

[6] Note that the parameter space for $\tilde{\phi}$ is the compact set $\tilde{\Phi}$ and not the entire stationarity region of a (causal) linear AR(p) model. Asymptotically, also the OLS estimator can be used.

conditions (ii) and (iii) in Assumption 1, A can be any compact subset of $\{(\alpha_0, \alpha_1, \ldots, \alpha_m) \in \mathbb{R}^{m+1} : (\alpha_1, \ldots, \alpha_m) \neq (0, \ldots, 0)\}$ where $1 \leq m \leq p$. This ensures that the mixing weight α_t^L is not equal to a constant. For the verification of Assumption 2(ii), see Appendix B.

When $m = 0$, the mixing weight α_t^L (and hence $1-\alpha_t^L$) is constant and the LMAR model reduces to the MAR model of Wong and Li (2000). In this special case our testing problem requires different and more complicated analyses than in the 'real' LMAR case where $m \geq 1$ and $(\alpha_1, \ldots, \alpha_r) \neq (0, \ldots, 0)$ (we shall discuss this point more later). Therefore, the conditions $m \geq 1$ and $(\alpha_1, \ldots, \alpha_m) \neq (0, \ldots, 0)$ will be assumed in the sequel. A similar restriction is made by Jeffries (1998) in his (first-order) logistic mixture autoregressive model to facilitate the derivation of the LR test (see the hypotheses at the end of p. 95 and the following discussion, as well as the end of p. 110).

GMAR Example. The GMAR model exemplifies the setting with common coefficients by assuming that the intercept terms in the two regimes are the same (note that this still allows for different means in the two regimes). As will be discussed in more detail in Section 3.3.1, this assumption is partly due to the fact that otherwise the derivation of the LR test would become extremely complicated. Hence, in this example $\beta = \tilde{\phi}_0 \, (= \tilde{\varphi}_0)$, $\phi = (\tilde{\phi}_1, \ldots, \tilde{\phi}_p, \tilde{\sigma}_1^2)$, $\varphi = (\tilde{\varphi}_1, \ldots, \tilde{\varphi}_p, \tilde{\sigma}_2^2)$, $q_1 = 1$, and $q_2 = p + 1$. To satisfy Assumptions 1(ii) and (iii) the parameter space A of α can be any compact and convex subset of $(0, 1)$ (this also rules out the possibility that $\alpha = 0$ or $\alpha = 1$ discussed after Assumption 1). For the verification of Assumption 2(ii), see Appendix C.

It may be worth noting that there are cases where the mixing weight α_t^G is time invariant and equals α. If this happens the GMAR model reduces to the MAR model of Wong and Li (2000).[7] However, unlike in the case of the LMAR model this fact does not complicate the derivation of our test. The reason seems to be that in the GMAR model the reduction occurs only for certain values of the parameters $\tilde{\phi}$ and $\tilde{\varphi}$ whereas in the LMAR model it occurs for all values of $\tilde{\phi}$ and $\tilde{\varphi}$.

3.2 Reparameterized model

In standard testing problems the derivation of the asymptotic distribution of the LR test would rely on a quadratic expansion of the log-likelihood function $L_T(\alpha, \beta, \phi, \varphi) = \sum_{t=1}^{T} l_t(\alpha, \beta, \phi, \varphi)$; when the parameter α is not identified under the null hypothesis, the relevant derivatives in this expansion would be with respect to (β, ϕ, φ) for fixed values of $\alpha \in A$. In problems with a singular information matrix it turns out to be convenient to follow Rotnitzky et al. (2000) and Kasahara and Shimotsu (2012, 2015) and employ an appropriately reparameterized model.

The employed reparameterization is model specific and aims to have two conveniences. First, it transforms the null hypothesis $\phi = \varphi$ into a point null hypothesis where some components of the parameter vector are restricted to zero and the rest are left unrestricted. Second, and more importantly, it simplifies derivations in cases where the conventional quadratic expansion of the log-likelihood function breaks down because, under the null hypothesis, the scores of the parameters (β, ϕ, φ) are linearly dependent and, consequently, the (Fisher) information matrix is singular. As will be seen later, this is the case for the GMAR model of Kalliovirta et al. (2015) but not for the LMAR model of Wong and Li (2001).

General requirements for the reparameterization are described in the following assumption. Only the parameters restricted by the null hypothesis, ϕ and φ, are reparameterized. The examples in this and the following subsection illustrate how the reparameterization could be chosen.

[7] An example is when $p = 1$, $\tilde{\phi}_0 = \tilde{\varphi}_0 = 0$, and $\tilde{\sigma}_1^2/(1 - \tilde{\phi}_1^2) = \tilde{\sigma}_2^2/(1 - \tilde{\varphi}_1^2)$ where the last equality can hold even if $(\tilde{\phi}, \tilde{\sigma}_1^2)$ is different from $(\tilde{\varphi}, \tilde{\sigma}_2^2)$.

Assumption 3.

(i) *For every* $\alpha \in A$, *the mapping* $(\pi, \varpi) = \boldsymbol{\pi}_\alpha(\phi, \varphi)$ *from* $\Phi \times \Phi$ *to* Π_α *is one-to-one with* $\boldsymbol{\pi}_\alpha$ *and* $\boldsymbol{\pi}_\alpha^{-1}$ *continuous.*

(ii) *For every* $\alpha \in A$, $\boldsymbol{\pi}_\alpha((\Phi \times \Phi)^0) = \Phi \times \{0\}$ *and* $\boldsymbol{\pi}_\alpha(\phi^*, \phi^*) = (\pi^*, 0) := (\phi^*, 0)$.

(iii) $(\hat{\beta}_{T\alpha}, \hat{\pi}_{T\alpha}, \hat{\varpi}_{T\alpha}) = (\beta^*, \pi^*, 0) + o_{p\alpha}(1)$, *where* $(\hat{\pi}_{T\alpha}, \hat{\varpi}_{T\alpha}) := \boldsymbol{\pi}_\alpha(\hat{\phi}_{T\alpha}, \hat{\varphi}_{T\alpha})$.

We sometimes refer to the reparameterization described in Assumption 3 as the 'π-parameterization' and the original reparameterization as the 'ϕ-parameterization'. Note that the transformed parameters π and ϖ generally depend on α but, for brevity, we suppress this dependence from the notation. The parameter space of (π, ϖ) also depends on α and is given, for any $\alpha \in A$, by

$$\Pi_\alpha = \{(\pi, \varpi) \in \mathbb{R}^{2q_2} : (\pi, \varpi) = \boldsymbol{\pi}_\alpha(\phi, \varphi) \text{ for some } (\phi, \varphi) \in \Phi \times \Phi\}.$$

By Assumption 3(ii), the null hypothesis $\phi = \varphi$ can be equivalently written as $\varpi = 0$ or, more precisely, as

$$H_0 : \pi \in \Phi,\ \varpi = 0,\ \alpha \in A,\ \beta \in B \qquad \text{vs.} \qquad H_1 : (\pi, \varpi) \in \Pi_\alpha \setminus (\Phi \times \{0\}),\ \alpha \in A,\ \beta \in B.$$

Note that under H_0, the parameters β and π are identified, but α is not. As for Assumption 3(iii), it is a high level condition similar to Assumption 2(ii) from which it can be derived with appropriate additional assumptions. A simple Lipschitz condition similar to Andrews (1992, Assumption SE-1(b)), given in Lemma A.1 in Appendix A, is one possibility.

To develop further notation, partition $\boldsymbol{\pi}_\alpha^{-1}(\pi, \varpi)$ into two q_2–dimensional components as $\boldsymbol{\pi}_\alpha^{-1}(\pi, \varpi) = (\boldsymbol{\pi}_{\alpha,1}^{-1}(\pi, \varpi), \boldsymbol{\pi}_{\alpha,2}^{-1}(\pi, \varpi))$, and define

$$f_{2,t}^{\pi}(\alpha, \beta, \pi, \varpi) := f_2(y_t \mid \boldsymbol{y}_{t-1}; \alpha, \beta, \phi, \varphi) = \alpha_t^{\pi}(\alpha, \beta, \pi, \varpi) f_t(\beta, \boldsymbol{\pi}_{\alpha,1}^{-1}(\pi, \varpi)) + (1 - \alpha_t^{\pi}(\alpha, \beta, \pi, \varpi)) f_t(\beta, \boldsymbol{\pi}_{\alpha,2}^{-1}(\pi, \varpi)),$$

where $\alpha_t^{\pi}(\alpha, \beta, \pi, \varpi) = \alpha_t(\alpha, \beta, \boldsymbol{\pi}_{\alpha,1}^{-1}(\pi, \varpi), \boldsymbol{\pi}_{\alpha,2}^{-1}(\pi, \varpi))$ and the function $f_t(\cdot)$ is as in (2). The log-likelihood function of the reparameterized model can now be expressed as

$$L_T^{\pi}(\alpha, \beta, \pi, \varpi) = \sum_{t=1}^{T} l_t^{\pi}(\alpha, \beta, \pi, \varpi), \tag{8}$$

where $l_t^{\pi}(\alpha, \beta, \pi, \varpi) = \log[f_{2,t}^{\pi}(\alpha, \beta, \pi, \varpi)]$, and in the π–parameterization equation (7) reads as

$$LR_T(\alpha) = 2[L_T^{\pi}(\alpha, \hat{\beta}_{T\alpha}, \hat{\pi}_{T\alpha}, \hat{\varpi}_{T\alpha}) - L_T^{\pi}(\alpha, \beta^*, \phi^*, 0)] - 2[L_T^{0}(\hat{\tilde{\phi}}_T) - L_T^{0}(\tilde{\phi}^*)] + o_{p\alpha}(1). \tag{9}$$

3.2.1 Examples (continued)

LMAR Example. The reparameterization we employ in the LMAR model is

$$(\pi, \varpi) = \boldsymbol{\pi}_\alpha(\phi, \varphi) = (\phi, \phi - \varphi) \text{ so that } (\phi, \varphi) = \boldsymbol{\pi}_\alpha^{-1}(\pi, \varpi) = (\pi, \pi - \varpi).$$

Note that in this case the reparameterization (via $\boldsymbol{\pi}_\alpha(\cdot, \cdot)$) does not depend on α, and the same is true for the parameter space of (π, ϖ). Verification of Assumption 3 is straightforward using Lemma A.1 (for details, see Appendix B). In the LMAR case, the only benefit of the reparameterization is to transform the null hypothesis into a point null hypothesis.

GMAR Example. In the GMAR model our reparameterization is obtained by setting, for any fixed $\alpha \in A$,

$$(\pi, \varpi) = \boldsymbol{\pi}_\alpha(\phi, \varphi) = (\alpha\phi + (1-\alpha)\varphi, \phi - \varphi) \text{ so that } (\phi, \varphi) = \boldsymbol{\pi}_\alpha^{-1}(\pi, \varpi) = (\pi + (1-\alpha)\varpi, \pi - \alpha\varpi).$$

Verification of Assumption 3 is again straightforward using Lemma A.1 (for details, see Appendix C). In the GMAR model, simplifying the null hypothesis is not the only benefit of the reparameterization, as will be discussed next.

As discussed before Assumption 3, the relevant derivatives when expanding $L_T(\alpha, \beta, \phi, \varphi)$ are with respect to (β, ϕ, φ) and, in the GMAR case, these derivatives are linearly dependent under the null hypothesis. To see this and how the reparameterization affects this feature, note first that straightforward differentiation yields

$$\nabla_{(\beta,\phi,\varphi)} l_t(\alpha, \beta, \phi, \phi) = \left(\alpha \frac{\nabla_\beta f_t(\beta, \phi)}{f_t(\beta, \phi)}, \alpha \frac{\nabla_\phi f_t(\beta, \phi)}{f_t(\beta, \phi)}, (1-\alpha) \frac{\nabla_\phi f_t(\beta, \phi)}{f_t(\beta, \phi)} \right),$$

where the null hypothesis $\phi = \varphi$ is imposed and ∇ denotes differentiation with respect to the indicated parameters. As $(f_t(\beta, \phi))^{-1} \nabla_{(\beta,\phi)} f_t(\beta, \phi)$ is the score vector obtained from a linear Gaussian AR(p) model, it is clear that the covariance matrix of the $(2p+3)$–dimensional vector $\nabla_{(\beta,\phi,\varphi)} l_t(\alpha, \beta, \phi, \phi)$, and hence the (Fisher) information matrix, is singular with rank $p+2$. In contrast to the above, in the π–parameterization the score vector is given by (see Supplementary Appendix C)

$$\nabla_{(\beta,\pi,\varpi)} l_t^\pi(\alpha, \beta, \pi, 0) = \left(\frac{\nabla_\beta f_t(\beta, \pi)}{f_t(\beta, \pi)}, \frac{\nabla_\pi f_t(\beta, \pi)}{f_t(\beta, \pi)}, \mathbf{0}_{p+1} \right)$$

when the null hypothesis $\varpi = 0$ is imposed. Now the score of ϖ is identically zero so that the reparameterization simplifies linear dependencies of the scores which turns out to be very useful in subsequent asymptotic analyses.

3.3 Quadratic expansion of the (reparameterized) log-likelihood function

As alluded to above, in standard testing problems the asymptotic analysis of a LR test statistic is based on a second order Taylor expansion of the (average) log-likelihood function around the true parameter value. An essential assumption here is positive definiteness of the (limiting) information matrix but, as illustrated in the previous section, this assumption does not necessarily hold in our testing problem due to linear dependencies among the derivatives of the log-likelihood function. As in Rotnitzky et al. (2000), Zhu and Zhang (2006), and Kasahara and Shimotsu (2012, 2015), we therefore consider a quadratic expansion of the log-likelihood function that is not based on a second order Taylor expansion but (possibly) on a higher order Taylor expansion. The need for higher-order derivatives is illustrated by the GMAR example: as the score of ϖ is identically zero, the second derivative (which turns out to be linearly independent of the score of (β, π)) now provides the first (nontrivial) local approximation for ϖ.

The following assumption ensures that the (reparameterized) log-likelihood function (8) is (at least) twice continuously differentiable.

Assumption 4. *For some integer $k \geq 2$, and for every fixed $\alpha \in A$, the functions $\alpha_t(\alpha, \beta, \phi, \varphi)$ and $\boldsymbol{\pi}_\alpha^{-1}(\pi, \varpi)$ are k times continuously differentiable (with respect to (β, ϕ, φ) and (π, ϖ) in the interior of $B \times \Phi \times \Phi$ and Π_α, respectively).*

In our general framework the reparameterized log-likelihood function is assumed to have, for each $\alpha \in A$, a quadratic expansion in a transformed parameter vector $\boldsymbol{\theta}(\alpha, \beta, \pi, \varpi)$ around $(\beta^*, \pi^*, 0)$ given by

$$
\begin{aligned}
&L_T^{\pi}(\alpha,\beta,\pi,\varpi) - L_T^{\pi}(\alpha,\beta^*,\pi^*,0) \\
&= (T^{-1/2}S_{T\alpha})'[T^{1/2}\boldsymbol{\theta}(\alpha,\beta,\pi,\varpi)] - \frac{1}{2}[T^{1/2}\boldsymbol{\theta}(\alpha,\beta,\pi,\varpi)]'\mathcal{I}_\alpha[T^{1/2}\boldsymbol{\theta}(\alpha,\beta,\pi,\varpi)] + R_T(\alpha,\beta,\pi,\varpi). \quad (10)
\end{aligned}
$$

To illustrate this expansion, suppose the information matrix is positive definite so that the quantities on the right hand side are (typically) based on a second order Taylor expansion with $S_{T\alpha}$ and $\mathcal{I}_\alpha$ functions of $(\alpha,\beta^*,\pi^*,0)$. As already mentioned, this is the case for the LMAR model where (the following will be justified shortly) the parameter $\boldsymbol{\theta}(\alpha,\beta,\pi,\varpi)$ is independent of α and given by $(\pi-\pi^*,\varpi)$ and, for each $\alpha\in A$, $S_{T\alpha}$ is the score vector, $\mathcal{I}_\alpha$ is the (positive definite) Fisher information matrix, and $R_T(\alpha,\beta,\pi,\varpi)$ is a remainder term. As the notation indicates, these three terms depend on α, and in general they may involve partial derivatives of the log-likelihood function of order higher than two (this is the case for the GMAR model, as will be demonstrated shortly). Then it may also get more complicated to find the reparameterization of the previous subsection and the transformed parameter vector $\boldsymbol{\theta}(\alpha,\beta,\pi,\varpi)$, as the examples of Kasahara and Shimotsu (2015, 2017) and the discussion on the GMAR model below show; one possibility is to consider the iterative procedure discussed by Rotnitzky et al. (2000, Sections 4.4 and 4.5) (for a recent illuminating illustration of this approach, see Hallin and Ley (2014)).

Our next assumption provides further details on expansion (10). We use $\Rightarrow$ to signify weak convergence of a sequence of stochastic processes on a function space. In the assumption below, the weak convergence of interest is that of the process $S_{T\alpha}$ (indexed by $\alpha\in A$) to a process S_α. The two function spaces relevant in this paper are $\mathcal{B}(A,\mathbb{R}^k)$ and $\mathcal{C}(A,\mathbb{R}^k)$, the former is the space of all $\mathbb{R}^k$–valued bounded functions defined on (the compact set) A equipped with the uniform metric ($d(x,y)=\sup_{a\in A}\|x(a)-y(a)\|$), and the latter is the same but with the continuity of the functions (with respect to $\alpha\in A$) also assumed.

Assumption 5. *For each $\alpha\in A$, the log-likelihood function $L_T^{\pi}(\alpha,\beta,\pi,\varpi)$ has a quadratic expansion given in (10), where*

(i) *for each $\alpha\in A$, $\boldsymbol{\theta}(\alpha,\beta,\pi,\varpi)$ is a mapping from $B\times\Pi_\alpha$ to $\Theta_\alpha=\{\boldsymbol{\theta}\in\mathbb{R}^r:\boldsymbol{\theta}=\boldsymbol{\theta}(\alpha,\beta,\pi,\varpi)\text{ for some }(\beta,\pi,\varpi)\in B\times\Pi_\alpha\}$ such that (a) $\boldsymbol{\theta}(\alpha,\beta^*,\pi^*,0)=0$ and (b) for all $\epsilon>0$, $\inf_{\alpha\in A}\inf_{(\beta,\pi,\varpi)\in B\times\Pi_\alpha:\|(\beta,\pi,\varpi)-(\beta^*,\pi^*,0)\|\geq\epsilon}\|\boldsymbol{\theta}(\alpha,\beta,\pi,\varpi)\|\geq\delta_\epsilon$ for some $\delta_\epsilon>0$.*

(ii) *$S_{T\alpha}=\sum_{t=1}^{T}s_{t\alpha}$ is a sequence of $\mathbb{R}^r$–valued $\mathcal{F}_T$–measurable stochastic processes indexed by $\alpha\in A$; $S_{T\alpha}$ does not depend on (β,π,ϖ); $S_{T\alpha}$ has sample paths that are continuous as functions of α; the process $T^{-1/2}S_{T\alpha}$ obeys $T^{-1/2}S_{T\bullet}\Rightarrow S_\bullet$ for some mean zero $\mathbb{R}^r$-valued Gaussian process $\{S_\alpha:\alpha\in A\}$ that satisfies $E[S_\alpha S_\alpha']=E[s_{t\alpha}s_{t\alpha}']=\mathcal{I}_\alpha$ for all $\alpha\in A$ and has continuous sample paths (as functions of α) with probability 1.*

(iii) *$\mathcal{I}_\alpha$ is, for each $\alpha\in A$, a non-random symmetric $r\times r$ matrix (independent of (β,π,ϖ)); $\mathcal{I}_\alpha$ is continuous as a function of α and such that $0<\inf_{\alpha\in A}\lambda_{\min}(\mathcal{I}_\alpha)$, $\sup_{\alpha\in A}\lambda_{\max}(\mathcal{I}_\alpha)<\infty$.*

(iv) *$R_T(\alpha,\beta,\pi,\varpi)$ is a remainder term such that*

$$
\sup_{(\beta,\pi,\varpi)\in B\times\Pi_\alpha:\|(\beta,\pi,\varpi)-(\beta^*,\pi^*,0)\|\leq\gamma_T}\frac{|R_T(\alpha,\beta,\pi,\varpi)|}{(1+\|T^{1/2}\boldsymbol{\theta}(\alpha,\beta,\pi,\varpi)\|)^2}=o_{p\alpha}(1)
$$

for all sequences of (non-random) positive scalars $\{\gamma_T,\ T\geq1\}$ for which $\gamma_T\to0$ as $T\to\infty$.

Assumption 5(i) describes the transformed parameter $\boldsymbol{\theta}(\alpha, \beta, \pi, \varpi)$, with part (b) being an identification condition. Assumption 5(ii) is the main ingredient needed to derive the limiting distribution of our LR test whereas 5(iv) ensures that the remainder term $R_T(\alpha, \beta, \pi, \varpi)$ has no effect on the final result. Assumption 5(iii) imposes rather standard conditions on the counterpart of the information matrix.

As in Andrews (1999, 2001), Zhu and Zhang (2006), and Kasahara and Shimotsu (2012, 2015), for further developments it will be convenient to write the expansion (10) in an alternative form as

$$\begin{aligned} &L_T^{\pi}(\alpha, \beta, \pi, \varpi) - L_T^{\pi}(\alpha, \beta^*, \pi^*, 0) \\ &\quad = \frac{1}{2} Z'_{T\alpha} \mathcal{I}_{\alpha} Z_{T\alpha} - \frac{1}{2}\big[T^{1/2}\boldsymbol{\theta}(\alpha, \beta, \pi, \varpi) - Z_{T\alpha}\big]' \mathcal{I}_{\alpha} \big[T^{1/2}\boldsymbol{\theta}(\alpha, \beta, \pi, \varpi) - Z_{T\alpha}\big] + R_T(\alpha, \beta, \pi, \varpi), \end{aligned} \tag{11}$$

where $Z_{T\alpha} = \mathcal{I}_{\alpha}^{-1} T^{-1/2} S_{T\alpha}$. Assumptions 5(ii) and (iii) imply the following facts (that will be justified in the proof of Lemma 1 in Appendix A): $Z_{T\alpha}$ is $\mathcal{F}_T$–measurable, independent of (β, π, ϖ), continuous as a function of α with probability 1, and $Z_{T\bullet} \Rightarrow Z_{\bullet}$ where the mean zero $\mathbb{R}^r$–valued Gaussian process $Z_{\alpha} = \mathcal{I}_{\alpha}^{-1} S_{\alpha}$ satisfies $E[Z_{\alpha} Z'_{\alpha}] = \mathcal{I}_{\alpha}^{-1}$ for all $\alpha \in A$ and has continuous sample paths (as functions of α) with probability 1.

3.3.1 Examples (continued)

LMAR Example. For the LMAR model, expansion (10) (with the unnecessary β being dropped everywhere) is obtained from a standard second-order Taylor expansion. Specifically, for an arbitrary fixed $\alpha \in A$, a standard second-order Taylor expansion of $L_T^{\pi}(\alpha, \pi, \varpi) = \sum_{t=1}^{T} l_t^{\pi}(\alpha, \pi, \varpi)$ around $(\pi^*, 0)$ with respect to the parameters (π, ϖ) yields

$$\begin{aligned} L_T^{\pi}(\alpha, \pi, \varpi) - L_T^{\pi}(\alpha, \pi^*, 0) &= (\pi - \pi^*, \varpi)' \nabla_{(\pi,\varpi)} L_T^{\pi}(\alpha, \pi^*, 0) \\ &\quad + \frac{1}{2}(\pi - \pi^*, \varpi)' \nabla^2_{(\pi,\varpi)(\pi,\varpi)'} L_T^{\pi}(\alpha, \dot{\pi}, \dot{\varpi})(\pi - \pi^*, \varpi), \end{aligned} \tag{12}$$

where $(\dot{\pi}, \dot{\varpi})$ denotes a point between (π, ϖ) and $(\pi^*, 0)$, $\nabla_{(\pi,\varpi)} L_T^{\pi}(\alpha, \pi^*, 0) = \sum_{t=1}^{T} \nabla_{(\pi,\varpi)} l_t^{\pi}(\alpha, \pi^*, 0)$ and $\nabla^2_{(\pi,\varpi)(\pi,\varpi)'} L_T^{\pi}(\alpha, \pi, \varpi) = \sum_{t=1}^{T} \nabla^2_{(\pi,\varpi)(\pi,\varpi)'} l_t^{\pi}(\alpha, \pi, \varpi)$ (explicit expressions for the required derivatives are provided in Appendix B), and ∇ and ∇^2 denote first and second order differentiation with respect to the indicated parameters. Set $\boldsymbol{\theta}(\alpha, \pi, \varpi) = (\pi - \pi^*, \varpi) = (\theta, \vartheta)$ and note that the parameter space $\Theta_{\alpha} = \Theta$ is independent of α and contains the origin, corresponding to the true model, as an interior point. Then define the vector $S_{T\alpha}$ and the matrix $\mathcal{I}_{\alpha}$ as[8]

$$S_{T\alpha} = \nabla_{(\pi,\varpi)} L_T^{\pi}(\alpha, \pi^*, 0) = \sum_{t=1}^{T} \left(\frac{\nabla f_t(\pi^*)}{f_t(\pi^*)}, -(1 - \alpha_{1,t}^L(\alpha)) \frac{\nabla f_t(\pi^*)}{f_t(\pi^*)} \right), \tag{13}$$

$$\mathcal{I}_{\alpha} = \begin{bmatrix} E\left[\frac{\nabla f_t(\pi^*)}{f_t(\pi^*)} \frac{\nabla' f_t(\pi^*)}{f_t(\pi^*)}\right] & -E\left[(1 - \alpha_{1,t}^L(\alpha)) \frac{\nabla f_t(\pi^*)}{f_t(\pi^*)} \frac{\nabla' f_t(\pi^*)}{f_t(\pi^*)}\right] \\ -E\left[(1 - \alpha_{1,t}^L(\alpha)) \frac{\nabla f_t(\pi^*)}{f_t(\pi^*)} \frac{\nabla' f_t(\pi^*)}{f_t(\pi^*)}\right] & E\left[(1 - \alpha_{1,t}^L(\alpha))^2 \frac{\nabla f_t(\pi^*)}{f_t(\pi^*)} \frac{\nabla' f_t(\pi^*)}{f_t(\pi^*)}\right] \end{bmatrix}.$$

Adding and subtracting terms and reorganizing, expansion (12) can be written as

$$\begin{aligned} L_T^{\pi}(\alpha, \pi, \varpi) - L_T^{\pi}(\alpha, \pi^*, 0) &= (T^{-1/2} S_{T\alpha})' [T^{1/2}\boldsymbol{\theta}(\alpha, \pi, \varpi)] \\ &\quad - \frac{1}{2}[T^{1/2}\boldsymbol{\theta}(\alpha, \pi, \varpi)]' \mathcal{I}_{\alpha} [T^{1/2}\boldsymbol{\theta}(\alpha, \pi, \varpi)] + R_T(\alpha, \pi, \varpi), \end{aligned} \tag{14}$$

[8]In what follows, $\nabla f_t(\cdot)$ denotes differentiation of $f_t(\cdot)$ in (3) with respect to $\tilde{\phi} = (\tilde{\phi}_0, \tilde{\phi}_1, \ldots, \tilde{\phi}_p, \tilde{\sigma}_1^2)$.

with the remainder term

$$R_T(\alpha,\pi,\varpi) = \frac{1}{2}[T^{1/2}\boldsymbol{\theta}(\alpha,\pi,\varpi)]'[T^{-1}\nabla^2_{(\pi,\varpi)(\pi,\varpi)'}L_T^{\pi}(\alpha,\dot{\pi},\dot{\varpi}) - (-\mathcal{I}_\alpha)][T^{1/2}\boldsymbol{\theta}(\alpha,\pi,\varpi)]. \tag{15}$$

These equations yield the expansion (10) in the case of the LMAR model. For details of verifying Assumptions 4 and 5, we refer to Appendix B.

As mentioned in the LMAR example of Section 3.1.1, the treatment of the special case where the mixing weight α_t^L is constant is more complicated than that of the 'real' LMAR case. Indeed, replacing the mixing weight α_t^L by a constant in the preceding expression of the score vector $S_{T\alpha}$ immediately shows that the second-order Taylor expansion (14) breaks down because, contrary to Assumption 5(iii), the components of $S_{T\alpha}$ are not linearly independent and, consequently, the Fisher information matrix $\mathcal{I}_\alpha$ is singular. Thus, a higher order Taylor expansion is needed to analyze the LR test statistic.

To give an idea of how one could proceed, we first note that the partial derivatives of the log-likelihood function behave in the same way as their counterparts in Kasahara and Shimotsu (2015) where mixtures of normal regression models (with constant mixing weights) are considered (see particularly the discussion following their Proposition 1). This is due to the fact that in the special case of constant mixing weights the LMAR model is obtained from the model considered in Kasahara and Shimotsu (2015) by replacing the exogenous regressors therein by lagged values of y_t. Thus, the arguments employed in that paper could be used to obtain the asymptotic distribution of the LR test statistic. Instead of a conventional second-order Taylor expansion this would require a more complicated reparameterization and an expansion based on partial derivatives of the log-likelihood function up to order eight. As most of the details appear very similar to those in Kasahara and Shimotsu (2015) we have preferred not to pursue this matter in this paper.

The preceding discussion means that, in the case of the LMAR model, time varying mixing weights are beneficial when the purpose is to derive a LR test for the adequacy of a single-regime model. A similar observation was made already by Jeffries (1998, p. 80). However, this does not happen in all mixture autoregressive models with time varying mixing weights, as the following discussion on the GMAR model demonstrates.

GMAR Example. As alluded to earlier, in the case of the GMAR model the expansion (10) cannot be based on a second order Taylor expansion of the log-likelihood function. A higher order expansion is required, and similarly to Kasahara and Shimotsu (2012) the appropriate order turns out to be the fourth one with the elements of $\nabla_\beta l_t^{\pi}(\alpha,\beta^*,\pi^*,0)$ and $\nabla_\pi l_t^{\pi}(\alpha,\beta^*,\pi^*,0)$ and the distinctive elements of $\nabla_{\varpi\varpi'} l_t^{\pi}(\alpha,\beta^*,\pi^*,0)$ (suitably normalized) used to define the vector $S_{T\alpha}$. In Appendix C we present, for an arbitrary fixed $\alpha \in A$, the explicit form of a standard fourth-order Taylor expansion of $L_T^{\pi}(\alpha,\beta,\pi,\varpi) = \sum_{t=1}^{T} l_t^{\pi}(\alpha,\beta,\pi,\varpi)$ around $(\beta^*,\pi^*,0)$ with respect to the parameters (β,π,ϖ). Therein we also demonstrate that this fourth-order Taylor expansion can be written as a quadratic expansion of the form (10) (or (11)) with the different quantities appearing therein defined as follows.

Define the vector $\boldsymbol{\theta}(\alpha,\beta,\pi,\varpi)$ in (10) as

$$\boldsymbol{\theta}(\alpha,\beta,\pi,\varpi) = \begin{bmatrix} \theta(\alpha,\beta,\pi,\varpi) \\ \vartheta(\alpha,\beta,\pi,\varpi) \end{bmatrix} = \begin{bmatrix} \beta-\beta^* \\ \pi-\pi^* \\ \alpha(1-\alpha)v(\varpi) \end{bmatrix},$$

where $\theta(\alpha,\beta,\pi,\varpi)$ is $(q_1+q_2)\times 1$ and $\vartheta(\alpha,\beta,\pi,\varpi)$ is $q_\vartheta \times 1$ with $q_\vartheta = q_2(q_2+1)/2$ (where $q_1 = 1$ and $q_2 = p+1$) and where the vector $v(\varpi)$ contains the unique elements of $\varpi\varpi'$, that is,

$$v(\varpi) = (\varpi_1^2,\ldots,\varpi_{q_2}^2,\varpi_1\varpi_2,\ldots,\varpi_1\varpi_{q_2},\varpi_2\varpi_3,\ldots,\varpi_{q_2-1}\varpi_{q_2})$$

(note that $v(\varpi)$ is just a re-ordering of $vech(\varpi\varpi'))$. The parameter space

$$\Theta_\alpha = \{\boldsymbol{\theta} = (\theta, \vartheta) \in \mathbb{R}^{q_1+q_2+q_\vartheta} : \theta = (\beta - \beta^*, \pi - \pi^*), \vartheta = \alpha(1-\alpha)v(\varpi) \text{ for some } (\beta, \pi, \varpi) \in B \times \Pi_\alpha\}$$

now depends on α and has the origin, corresponding to the true model, as a boundary point (due to the particular shape of the range of $\vartheta = \alpha(1-\alpha)v(\varpi)$); both of these features will complicate the subsequent analysis.

Next set

$$S_T\,(=S_{T\alpha}) = \sum_{t=1}^{T} \tilde{\nabla}_{\boldsymbol{\theta}} l_t^{\pi*} \qquad \text{where} \qquad \tilde{\nabla}_{\boldsymbol{\theta}} l_t^{\pi*} = (\tilde{\nabla}_\theta l_t^{\pi*}, \tilde{\nabla}_\vartheta l_t^{\pi*}) \qquad ((q_1+q_2+q_\vartheta)\times 1) \tag{16}$$

with the component vectors $\tilde{\nabla}_\theta l_t^{\pi*}$ and $\tilde{\nabla}_\vartheta l_t^{\pi*}$ given by

$$\begin{aligned}
\tilde{\nabla}_\theta l_t^{\pi*} &= (\nabla_\beta l_t^\pi(\alpha, \beta^*, \pi^*, 0), \nabla_\pi l_t^\pi(\alpha, \beta^*, \pi^*, 0)), \\
\tilde{\nabla}_\vartheta l_t^{\pi*} &= \big(c_{11}\nabla^2_{\varpi_1\varpi_1} l_t^\pi(\alpha, \beta^*, \pi^*, 0), \ldots, c_{q_2q_2}\nabla^2_{\varpi_{q_2}\varpi_{q_2}} l_t^\pi(\alpha, \beta^*, \pi^*, 0), \\
&\qquad c_{12}\nabla^2_{\varpi_1\varpi_2} l_t^\pi(\alpha, \beta^*, \pi^*, 0), \ldots, c_{q_2-1,q_2}\nabla^2_{\varpi_{q_2-1}\varpi_{q_2}} l_t^\pi(\alpha, \beta^*, \pi^*, 0)\big)/(\alpha(1-\alpha)),
\end{aligned} \tag{17}$$

where $c_{ij} = 1/2$ if $i = j$ and $c_{ij} = 1$ if $i \neq j$. Explicit expressions for $\tilde{\nabla}_\theta l_t^{\pi*}$ and $\tilde{\nabla}_\vartheta l_t^{\pi*}$ can be found in Appendix C, and from them it can be seen that S_T depends on (β^*, π^*) only and not on $(\alpha, \beta, \pi, \varpi)$. The same is true for the matrix $\mathcal{I}\,(=\mathcal{I}_\alpha)$ $((q_1+q_2+q_\vartheta)\times(q_1+q_2+q_\vartheta))$ whose expression is also given in Appendix C. Finally, an explicit expression of the remainder term $R_T(\alpha, \beta, \pi, \varpi)$ is given in Appendix C. For the verification of Assumptions 4 and 5, see Appendix C.

In the GMAR example we have assumed that the intercept terms $\tilde{\phi}_0$ and $\tilde{\varphi}_0$ in the two regimes are the same. We are now in a position to describe the difficulties that allowing for $\tilde{\phi}_0 \neq \tilde{\varphi}_0$ (and, hence, dropping β) would entail. In this case, the additional parameter $\tilde{\varphi}_0$ would correspond to ϖ_1, the first component of ϖ. As in Section 3.2.1, it would again be the case that $\nabla_{\varpi_1} l_t^\pi(\alpha, \pi^*, 0) = 0$, leading us to consider second derivatives. But now, due to the properties of the Gaussian distribution, it would be the case that $\nabla^2_{\varpi_1\varpi_1} l_t^\pi(\alpha, \pi^*, 0)$ is linearly dependent with the components of $\nabla_\pi l_t^\pi(\alpha, \pi^*, 0)$, making it unsuitable to serve as a component of S_T. A reparameterization more complicated than that used in Section 3.2.1 would be needed, with the aim of obtaining $\nabla^2_{\varpi_1\varpi_1} l_t^\pi(\alpha, \pi^*, 0) = 0$ and, instead of $\nabla^2_{\varpi_1\varpi_1} l_t^\pi(\alpha, \pi^*, 0)$, using $\nabla^3_{\varpi_1\varpi_1\varpi_1} l_t^\pi(\alpha, \pi^*, 0)$ or perhaps $\nabla^4_{\varpi_1\varpi_1\varpi_1\varpi_1} l_t^\pi(\alpha, \pi^*, 0)$ as the counterpart of the score of the parameter ϖ_1. It turns out that (restricting the discussion to the case $p = 1$ only) the third derivative is suitable when $\alpha \neq 1/2$ and $\tilde{\phi}_1 \neq -1/2$, but that fourth (or higher) order derivatives are needed when $\alpha = 1/2$ or $\tilde{\phi}_1 = -1/2$. Similar difficulties (involving situations comparable to the cases $\alpha = 1/2$ vs. $\alpha \neq 1/2$, but apparently not ones involving also a counterpart of $\tilde{\phi}_1$) were faced by Cho and White (2007, Sec. 2.3.3) and Kasahara and Shimotsu (2017, Sec. 6.2), whose analyses suggest that expanding the log-likelihood at least to the eighth order is required. As the required analysis gets excessively complicated, we have chosen to leave it for future research.

3.4 Asymptotic analysis of the quadratic expansion

We continue by analyzing the expansion (11) evaluated at $(\hat{\beta}_{T\alpha}, \hat{\pi}_{T\alpha}, \hat{\varpi}_{T\alpha})$. Previously, a similar analysis is provided by Andrews (2001) but his approach is not directly applicable in our setting. The reason for this is that in the quadratic expansion in (11) the dependence of the parameter $\boldsymbol{\theta}(\alpha, \beta, \pi, \varpi)$ and its parameter space Θ_α on the nuisance parameter α is not compatible with the formulation of Andrews (2001, eqn (3.3)). The results of Zhu and Zhang (2006) probably cover our case, but instead of trying to verify the assumptions employed by these authors we prove the needed results by adapting the arguments used in Andrews (1999, 2001) and Zhu and Zhang (2006) to our setting. We proceed in several steps.

Asymptotic insignificance of the remainder term. We first establish that the remainder term $R_T(\alpha,\beta,\pi,\varpi)$, when evaluated at $(\hat{\beta}_{T\alpha},\hat{\pi}_{T\alpha},\hat{\varpi}_{T\alpha})$, has no effect on the asymptotic distribution of the quadratic expansion. A crucial ingredient in showing this is showing that the transformed parameter vector $\boldsymbol{\theta}(\alpha,\hat{\beta}_{T\alpha},\hat{\pi}_{T\alpha},\hat{\varpi}_{T\alpha})$ is root-T consistent in the sense that $\|T^{1/2}\boldsymbol{\theta}(\alpha,\hat{\beta}_{T\alpha},\hat{\pi}_{T\alpha},\hat{\varpi}_{T\alpha})\| = O_{p\alpha}(1)$. This, together with part (iv) of Assumption 5 allows us to obtain the result $R_T(\alpha,\hat{\beta}_{T\alpha},\hat{\pi}_{T\alpha},\hat{\varpi}_{T\alpha}) = o_{p\alpha}(1)$. We collect these results in the following lemma.

Lemma 1. *If Assumptions 1–5 hold*[9]*, then (i)* $\|T^{1/2}\boldsymbol{\theta}(\alpha,\hat{\beta}_{T\alpha},\hat{\pi}_{T\alpha},\hat{\varpi}_{T\alpha})\| = O_{p\alpha}(1)$*,*
(ii) $R_T(\alpha,\hat{\beta}_{T\alpha},\hat{\pi}_{T\alpha},\hat{\varpi}_{T\alpha}) = o_{p\alpha}(1)$*, and (iii)*

$$\begin{aligned}&L_T^{\pi}(\alpha,\hat{\beta}_{T\alpha},\hat{\pi}_{T\alpha},\hat{\varpi}_{T\alpha}) - L_T^{\pi}(\alpha,\beta^*,\pi^*,0)\\ &= \frac{1}{2}Z_{T\alpha}'\mathcal{I}_\alpha Z_{T\alpha} - \frac{1}{2}\big[T^{1/2}\boldsymbol{\theta}(\alpha,\hat{\beta}_{T\alpha},\hat{\pi}_{T\alpha},\hat{\varpi}_{T\alpha}) - Z_{T\alpha}\big]'\mathcal{I}_\alpha\big[T^{1/2}\boldsymbol{\theta}(\alpha,\hat{\beta}_{T\alpha},\hat{\pi}_{T\alpha},\hat{\varpi}_{T\alpha}) - Z_{T\alpha}\big] + o_{p\alpha}(1).\end{aligned} \tag{18}$$

Note that assertion (iii) of Lemma 1 is analogous to Andrews (1999, Theorem 2b).

Maximization of the likelihood vs. minimization of a related quadratic form. The first two terms on the right hand side of (11) provide an approximation to the (reparameterized and centered) log-likelihood function $L_T^{\pi}(\alpha,\beta,\pi,\varpi) - L_T^{\pi}(\alpha,\beta^*,\pi^*,0)$ evaluated at the (approximate unrestricted reparameterized) ML estimator $(\hat{\beta}_{T\alpha},\hat{\pi}_{T\alpha},\hat{\varpi}_{T\alpha})$ in equation (18). For later developments it would be convenient if the ML estimator on the right hand side of (18) could be replaced by an (approximate) minimizer of the quadratic form $[T^{1/2}\boldsymbol{\theta}(\alpha,\beta,\pi,\varpi) - Z_{T\alpha}]'\mathcal{I}_\alpha[T^{1/2}\boldsymbol{\theta}(\alpha,\beta,\pi,\varpi) - Z_{T\alpha}]$. In order to justify this replacement we first note that, by definition, (cf. Andrews (1999, eqn. (3.6)))

$$\inf_{(\beta,\pi,\varpi)\in B\times\Pi_\alpha}\big\{\big[T^{1/2}\boldsymbol{\theta}(\alpha,\beta,\pi,\varpi)-Z_{T\alpha}\big]'\mathcal{I}_\alpha\big[T^{1/2}\boldsymbol{\theta}(\alpha,\beta,\pi,\varpi)-Z_{T\alpha}\big]\big\} = \inf_{\boldsymbol{\lambda}\in\Theta_{\alpha,T}}\big\{(\boldsymbol{\lambda}-Z_{T\alpha})'\mathcal{I}_\alpha(\boldsymbol{\lambda}-Z_{T\alpha})\big\}$$

where, for each T,

$$\Theta_{\alpha,T} = \{\boldsymbol{\lambda}\in\mathbb{R}^r : \boldsymbol{\lambda} = T^{1/2}\boldsymbol{\theta} \text{ for some } \boldsymbol{\theta}\in\Theta_\alpha\}$$

and Θ_α is as defined in Assumption 5(i). Next, for each $\alpha\in A$, let $\hat{\boldsymbol{\lambda}}_{T\alpha q} = T^{1/2}\boldsymbol{\theta}(\alpha,\hat{\beta}_{T\alpha q},\hat{\pi}_{T\alpha q},\hat{\varpi}_{T\alpha q})$ (with the additional 'q' in the subscripts referring to quadratic form) denote an approximate minimizer of $(\boldsymbol{\lambda}-Z_{T\alpha})'\mathcal{I}_\alpha(\boldsymbol{\lambda}-Z_{T\alpha})$ over $\Theta_{\alpha,T}$, that is, (cf. Zhu and Zhang (2006, eqn. (12)))

$$(\hat{\boldsymbol{\lambda}}_{T\alpha q}-Z_{T\alpha})'\mathcal{I}_\alpha(\hat{\boldsymbol{\lambda}}_{T\alpha q}-Z_{T\alpha}) = \inf_{\boldsymbol{\lambda}\in\Theta_{\alpha,T}}\big\{(\boldsymbol{\lambda}-Z_{T\alpha})'\mathcal{I}_\alpha(\boldsymbol{\lambda}-Z_{T\alpha})\big\} + o_{p\alpha}(1). \tag{19}$$

Now we can state the following lemma justifying the discussed replacement of $(\hat{\beta}_{T\alpha},\hat{\pi}_{T\alpha},\hat{\varpi}_{T\alpha})$ with $\hat{\boldsymbol{\lambda}}_{T\alpha q}$.

Lemma 2. *If Assumptions 1–5 hold, then*

$$L_T^{\pi}(\alpha,\hat{\beta}_{T\alpha},\hat{\pi}_{T\alpha},\hat{\varpi}_{T\alpha}) - L_T^{\pi}(\alpha,\beta^*,\pi^*,0) = \frac{1}{2}Z_{T\alpha}'\mathcal{I}_\alpha Z_{T\alpha} - \frac{1}{2}(\hat{\boldsymbol{\lambda}}_{T\alpha q}-Z_{T\alpha})'\mathcal{I}_\alpha(\hat{\boldsymbol{\lambda}}_{T\alpha q}-Z_{T\alpha}) + o_{p\alpha}(1).$$

Approximating the parameter space with a cone. In the previous subsection the quadratic form $(\boldsymbol{\lambda}-Z_{T\alpha})'\mathcal{I}_\alpha(\boldsymbol{\lambda}-Z_{T\alpha})$ was minimized over the set $\Theta_{\alpha,T}$ which can be complicated and hence

[9] Here and in what follows, a subset of the listed assumptions would sometimes suffice for the stated results.

difficult to use. Therefore we next show that the quadratic form $(\boldsymbol{\lambda} - Z_{T\alpha})'\mathcal{I}_\alpha(\boldsymbol{\lambda} - Z_{T\alpha})$ can instead be minimized over a simpler set, and to this end we first introduce some terminology.

We say that a collection of sets $\{\Gamma_\alpha,\ \alpha \in A\}$ (where for each $\alpha \in A$, $\Gamma_\alpha \subset \mathbb{R}^r$) is 'locally (at the origin) uniformly equal' to a set $\Lambda \subset \mathbb{R}^r$ if there exists a $\delta > 0$ such that $\Gamma_\alpha \cap (-\delta,\delta)^r = \Lambda \cap (-\delta,\delta)^r$ for all $\alpha \in A$. Note that '$\{\Gamma_\alpha,\ \alpha \in A\}$ is locally uniformly equal to Λ' implies that (i) 'for all $\alpha \in A$, Γ_α is locally equal to Λ in the sense of Andrews (1999, p. 1359)', but the reverse does not hold; and also that (ii) $\{\Gamma_\alpha,\ \alpha \in A\}$ is uniformly approximated by the set Λ in the sense of Zhu and Zhang (2006, Defn. 3). Finally, we say that a set $\Lambda \subset \mathbb{R}^r$ is a 'cone' if $\lambda \in \Lambda$ implies that $a\lambda \in \Lambda$ for all positive real scalars a.

Based on the preceding discussion we state the following assumption.

Assumption 6. *The collection of sets $\{\Theta_\alpha,\ \alpha \in A\}$ is locally uniformly equal to a cone Λ ($\subset \mathbb{R}^r$).*

Note that by Assumption 5(i)(a), $0 \in \Theta_\alpha$ for all $\alpha \in A$, so that the cone Λ in Assumption 6 necessarily contains 0 ($\in \mathbb{R}^r$). The cone Λ also does not depend on α. Now we can establish the following result.

Lemma 3. *If Assumptions 1–6 hold, then*

$$\inf_{\boldsymbol{\lambda}\in\Theta_{\alpha,T}} \left\{(\boldsymbol{\lambda} - Z_{T\alpha})'\mathcal{I}_\alpha(\boldsymbol{\lambda} - Z_{T\alpha})\right\} = \inf_{\boldsymbol{\lambda}\in\Lambda} \left\{(\boldsymbol{\lambda} - Z_{T\alpha})'\mathcal{I}_\alpha(\boldsymbol{\lambda} - Z_{T\alpha})\right\} + o_{p\alpha}(1).$$

Describing the limiting random variable. From Lemmas 2 and 3 and the definition of $\hat{\boldsymbol{\lambda}}_{T\alpha q}$ we can now conclude that

$$2[L_T^\pi(\alpha, \hat{\beta}_{T\alpha}, \hat{\pi}_{T\alpha}, \hat{\varpi}_{T\alpha}) - L_T^\pi(\alpha, \beta^*, \pi^*, 0)] = Z_{T\alpha}'\mathcal{I}_\alpha Z_{T\alpha} - \inf_{\boldsymbol{\lambda}\in\Lambda} \left\{(\boldsymbol{\lambda} - Z_{T\alpha})'\mathcal{I}_\alpha(\boldsymbol{\lambda} - Z_{T\alpha})\right\} + o_{p\alpha}(1). \tag{20}$$

The assumed weak convergence of $S_{T\alpha}$ (and hence that of $Z_\alpha = \mathcal{I}_\alpha^{-1}S_\alpha$) allows us to derive the weak limit of this random process described in the following lemma.

Lemma 4. *If Assumptions 1–6 hold, then*

$$2[L_T^\pi(\bullet, \hat{\beta}_{T\bullet}, \hat{\pi}_{T\bullet}, \hat{\varpi}_{T\bullet}) - L_T^\pi(\bullet, \beta^*, \pi^*, 0)] \Rightarrow Z_\bullet'\mathcal{I}_\bullet Z_\bullet - \inf_{\boldsymbol{\lambda}\in\Lambda} \left\{(\boldsymbol{\lambda} - Z_\bullet)'\mathcal{I}_\bullet(\boldsymbol{\lambda} - Z_\bullet)\right\}.$$

The limiting random process in Lemma 4 can be written in a somewhat simpler form (cf. Andrews (1999, Thm 4) and Andrews (2001, Thm 2)). The motivation for this comes from the fact that in our applications $\boldsymbol{\theta}(\alpha,\beta,\pi,\varpi)$ can be decomposed into two parts as $\boldsymbol{\theta}(\alpha,\beta,\pi,\varpi) = (\theta(\alpha,\beta,\pi,\varpi), \vartheta(\alpha,\beta,\pi,\varpi))$ with $\theta(\alpha,\beta,\pi,\varpi) \in \mathbb{R}^{q_\theta}$ and $\vartheta(\alpha,\beta,\pi,\varpi) \in \mathbb{R}^{q_\vartheta}$ (with $q_\theta = q_1 + q_2$ and $r = q_\theta + q_\vartheta$) such that (i) the values of $\theta(\alpha,\beta,\pi,\varpi)$ are not restricted by the null hypothesis and do not lie on the boundary of the parameter space and (ii) the values of $\vartheta(\alpha,\beta,\pi,\varpi)$ are restricted by the null hypothesis and potentially lie on the boundary of the parameter space. Specifically, we assume the following.

Assumption 7. *The cone Λ of Assumption 6 satisfies $\Lambda = \mathbb{R}^{q_\theta} \times \Lambda_\vartheta$ with Λ_ϑ a cone in $\mathbb{R}^{q_\vartheta}$.*

Partition S_α, Z_α, $\boldsymbol{\lambda}$, and $\mathcal{I}_\alpha$ conformably with the partition $\boldsymbol{\theta}(\alpha,\beta,\pi,\varpi) = (\theta(\alpha,\beta,\pi,\varpi), \vartheta(\alpha,\beta,\pi,\varpi))$ as

$$S_\alpha = \begin{bmatrix} S_{\theta\alpha} \\ S_{\vartheta\alpha} \end{bmatrix},\quad Z_\alpha = \begin{bmatrix} Z_{\theta\alpha} \\ Z_{\vartheta\alpha} \end{bmatrix},\quad \boldsymbol{\lambda} = \begin{bmatrix} \boldsymbol{\lambda}_\theta \\ \boldsymbol{\lambda}_\vartheta \end{bmatrix},\quad \mathcal{I}_\alpha = \begin{bmatrix} \mathcal{I}_{\theta\theta\alpha} & \mathcal{I}_{\theta\vartheta\alpha} \\ \mathcal{I}_{\vartheta\theta\alpha} & \mathcal{I}_{\vartheta\vartheta\alpha} \end{bmatrix}$$

and let $(\mathcal{I}_\alpha^{-1})_{\vartheta\vartheta}$ denote the $(q_\vartheta \times q_\vartheta)$ bottom right block of $\mathcal{I}_\alpha^{-1}$. Assumption 7 together with properties of partitioned matrices yields the following result.

Lemma 5. *If Assumptions 1–7 hold, then*

$$
\begin{aligned}
&Z'_\alpha \mathcal{I}_\alpha Z_\alpha - \inf_{\boldsymbol{\lambda}\in\Lambda} \left\{(\boldsymbol{\lambda} - Z_\alpha)'\mathcal{I}_\alpha(\boldsymbol{\lambda} - Z_\alpha)\right\} \\
&= Z'_{\vartheta\alpha}(\mathcal{I}_\alpha^{-1})^{-1}_{\vartheta\vartheta} Z_{\vartheta\alpha} - \inf_{\boldsymbol{\lambda}_\vartheta\in\Lambda_\vartheta} \left\{(\boldsymbol{\lambda}_\vartheta - Z_{\vartheta\alpha})'(\mathcal{I}_\alpha^{-1})^{-1}_{\vartheta\vartheta}(\boldsymbol{\lambda}_\vartheta - Z_{\vartheta\alpha})\right\} + S'_{\theta\alpha}\mathcal{I}^{-1}_{\theta\theta\alpha}S_{\theta\alpha}.
\end{aligned}
$$

Explicit expressions for $(\mathcal{I}_\alpha^{-1})_{\vartheta\vartheta}$ and $Z_{\vartheta\alpha}$ in terms of S_α and $\mathcal{I}_\alpha$ are given in the proof of this lemma in Supplementary Appendix D.

3.5 The LR test statistic

3.5.1 Derivation of the test statistic

The previous subsection described the asymptotic behavior of $2[L_T^\pi(\alpha, \hat{\beta}_{T\alpha}, \hat{\pi}_{T\alpha}, \hat{\varpi}_{T\alpha}) - L_T^\pi(\alpha, \beta^*, \pi^*, 0)]$, the first term in the expression of $LR_T(\alpha)$ in (9). Now consider the second term, namely $2[L_T^0(\hat{\tilde{\phi}}_T) - L_T^0(\tilde{\phi}^*)]$, corresponding to the model restricted by the null hypothesis. Recall that $L_T^0(\tilde{\phi}) = \sum_{t=1}^T l_t^0(\tilde{\phi})$ with $l_t^0(\tilde{\phi}) = \log[f_t(\tilde{\phi})]$ so that $\nabla_{\tilde{\phi}} l_t^0(\tilde{\phi}^*) = (f_t(\tilde{\phi}^*))^{-1}\nabla f_t(\tilde{\phi}^*)$ with $\tilde{\phi}^*$ an interior point of $\tilde{\Phi}$. Denote the score vector and limiting information matrix by

$$
S_T^0 = \sum_{t=1}^T \frac{\nabla f_t(\tilde{\phi}^*)}{f_t(\tilde{\phi}^*)}, \quad \mathcal{I}^0 = E\left[\frac{\nabla f_t(\tilde{\phi}^*)}{f_t(\tilde{\phi}^*)}\frac{\nabla' f_t(\tilde{\phi}^*)}{f_t(\tilde{\phi}^*)}\right],
$$

respectively. For the following assumption, partition the process $S_{T\alpha}$ of Assumption 5 as $S_{T\alpha} = (S_{T\theta\alpha}, S_{T\vartheta\alpha})$ (with $S_{T\theta\alpha}$ q_θ–dimensional and $S_{T\vartheta\alpha}$ q_ϑ–dimensional). The following simplifying assumption, which holds in our examples (see the expressions of $S_{T\alpha}$ in (13) and (16)), allows us to obtain a neat expression for the likelihood ratio test statistic in Theorem 1 below.

Assumption 8. $S_{T\theta\alpha} = S_T^0$.

Together with the earlier assumptions, Assumption 8 implies that $T^{-1/2}S_{T\theta\alpha} = T^{-1/2}S_T^0 \xrightarrow{d} S^0$, a q_θ–dimensional Gaussian random vector with mean zero and covariance matrix $\mathcal{I}_{\theta\theta\alpha} = E[S^0 S^{0\prime}] = \mathcal{I}^0$. Standard likelihood theory now implies the following result.

Lemma 6. *If Assumptions 1–8 hold, then* $2[L_T^0(\hat{\tilde{\phi}}_T) - L_T^0(\tilde{\phi}^*)] \xrightarrow{d} S^{0\prime}(\mathcal{I}^0)^{-1}S^0$, *and the convergence is joint with that in Lemma 4.*

The preceding results, in particular Lemmas 4, 5, and 6, now yield the distribution of the LR test statistic in the following theorem.

Theorem 1. *If Assumptions 1–8 hold, then*

(i) $LR_T(\bullet) \Rightarrow Z'_{\vartheta\bullet}(\mathcal{I}_\bullet^{-1})^{-1}_{\vartheta\vartheta}Z_{\vartheta\bullet} - \inf_{\boldsymbol{\lambda}_\vartheta\in\Lambda_\vartheta}\{(\boldsymbol{\lambda}_\vartheta - Z_{\vartheta\bullet})'(\mathcal{I}_\bullet^{-1})^{-1}_{\vartheta\vartheta}(\boldsymbol{\lambda}_\vartheta - Z_{\vartheta\bullet})\}$, *and*

(ii) $LR_T = \sup_{\alpha\in A} LR_T(\alpha) \xrightarrow{d} \sup_{\alpha\in A}\left\{Z'_{\vartheta\alpha}(\mathcal{I}_\alpha^{-1})^{-1}_{\vartheta\vartheta}Z_{\vartheta\alpha} - \inf_{\boldsymbol{\lambda}_\vartheta\in\Lambda_\vartheta}\{(\boldsymbol{\lambda}_\vartheta - Z_{\vartheta\alpha})'(\mathcal{I}_\alpha^{-1})^{-1}_{\vartheta\vartheta}(\boldsymbol{\lambda}_\vartheta - Z_{\vartheta\alpha})\}\right\}$.

This completes the derivation of the LR test statistic. The asymptotic distribution is similar to that in Andrews (2001, Thm 4). As we next discuss, this distribution simplifies in both the LMAR and the GMAR examples.

3.5.2 Examples (continued)

LMAR Example. As was noted in Section 3.3.1, the LMAR case is rather standard in the sense that a conventional second-order Taylor expansion with a nonsingular information matrix and with no parameters on the boundary was sufficient to study the LR test. The only nonstandard feature in this case is the presence of unidentified parameters under the null hypothesis. Validity of Assumptions 6–8 is easy to check (see Appendix B) with the cone Λ of Assumption 6 equal to $\mathbb{R}^r$. Thus the infimum in the distribution of the LR_T statistic in Theorem 1(ii) equals zero and the result therein simplifies to[10]

$$LR_T = \sup_{\alpha \in A} LR_T(\alpha) \xrightarrow{d} \sup_{\alpha \in A} \left\{ Z'_{\vartheta\alpha} (\mathcal{I}_\alpha^{-1})^{-1}_{\vartheta\vartheta} Z_{\vartheta\alpha} \right\}.$$

For every fixed $\alpha \in A$, the quantity $Z'_{\vartheta\alpha} (\mathcal{I}_\alpha^{-1})^{-1}_{\vartheta\vartheta} Z_{\vartheta\alpha}$ is a chi-squared random variable, so that the limiting distribution is a supremum of a chi-squared process similarly as in, for example, Davies (1987), Hansen (1996, Thm 1), and Andrews (2001, eqn. (5.7)).

GMAR Example. In Section 3.3.1 it was seen that in the GMAR example Z_α and $\mathcal{I}_\alpha$ do not depend on α. As the cone Λ of Assumption 6 does not depend on α either, the weak limit of $LR_T(\alpha)$ does not depend on α. Therefore the result of Theorem 1 (validity of Assumptions 6–8 is checked in Appendix C) simplifies to

$$LR_T = \sup_{\alpha \in A} LR_T(\alpha) \xrightarrow{d} Z'_\vartheta (\mathcal{I}^{-1})^{-1}_{\vartheta\vartheta} Z_\vartheta - \inf_{\boldsymbol{\lambda}_\vartheta \in \Lambda_\vartheta} \left\{ (\boldsymbol{\lambda}_\vartheta - Z_\vartheta)' (\mathcal{I}^{-1})^{-1}_{\vartheta\vartheta} (\boldsymbol{\lambda}_\vartheta - Z_\vartheta) \right\},$$

where the unnecessary α has been dropped from the notation. Here Z_ϑ follows an q_ϑ-variate Gaussian distribution with covariance matrix $(\mathcal{I}^{-1})_{\vartheta\vartheta}$, and the limiting distribution, which is sometimes referred to as the chi-bar-squared distribution, is similar to the one in Kasahara and Shimotsu (2012, Proposition 3c,d). Note that the cone $\Lambda_\vartheta = v(\mathbb{R}^{q_2})$ (see Appendix C) is not convex (in contrast to (at least most of) the examples in Andrews (2001), but similarly to Kasahara and Shimotsu (2012, Proposition 3c,d)) and the dimension of this cone, $q_\vartheta = q_2(q_2+1)/2$, may not be small either ($q_\vartheta = 3, 6, 10, \ldots$ for $q_2 = 2, 3, 4, \ldots$).

4 Simulation-based critical values and a Monte Carlo study

4.1 Simulating the asymptotic null distribution

Similarly to Hansen (1996) and Andrews (2001), the asymptotic null distribution of the LR statistic in Theorem 1 is typically application-specific and cannot be tabulated. Following these papers, we use simulation methods to obtain critical values of the asymptotic null distribution. The following procedure is based on Hansen (1996) and is analogous to the one used by Zhu and Zhang (2004, Sec 2.1) in a related mixture setting.[11]

Let A_G be some finite grid of α values in A. For each fixed $\alpha \in A_G$, let $\hat{s}_{t\alpha}$ signify an empirical counterpart of $s_{t\alpha}$ (see Assumption 5) where the unknown parameter $\tilde{\phi}^*$ (or (β^*, π^*)) is replaced by its consistent estimator under the null, $\hat{\tilde{\phi}}_T$. (The specific forms of $\hat{s}_{t\alpha}$ in the LMAR and GMAR

[10] This result could also be obtained from Andrews and Ploberger (1995, Sec 2.2, 2.4) as their assumptions 1–5 appear to be satisfied in the LMAR case.

[11] An alternative to this procedure is to use bootstrap. However, the validity of bootstrap in the presence of parameters on the boundary and singular information matrices is not clear (see, e.g., Andrews (2000)). Another reason for preferring the proposed simulation method is that repeated estimation of the mixture model under the alternative may be computationally rather demanding.

examples are provided in Appendices B and C, respectively.) Set $\hat{\mathcal{I}}_{T\alpha} = T^{-1}\sum_{t=1}^{T}\hat{s}_{t\alpha}\hat{s}'_{t\alpha}$. Now, for each $j = 1, \ldots, J$ (where J denotes the number of repetitions), do the following.

(i) Generate a sequence $\{v_{tj}\}_{t=1}^{T}$ of T i.i.d. $N(0,1)$ random variables.

(ii) For each $\alpha \in A_G$, set $\hat{S}^{j}_{T\alpha} = \sum_{t=1}^{T}\hat{s}_{t\alpha}v_{tj}$, $\hat{Z}^{j}_{T\alpha} = \hat{\mathcal{I}}^{-1}_{T\alpha}T^{-1/2}\hat{S}^{j}_{T\alpha}$, and (using similar partitioning notation as before)

$$\widehat{LR}^{j}_{T}(\alpha) = \hat{Z}^{j\prime}_{T\vartheta\alpha}(\hat{\mathcal{I}}^{-1}_{T\alpha})^{-1}_{\vartheta\vartheta}\hat{Z}^{j}_{T\vartheta\alpha} - \inf_{\boldsymbol{\lambda}_{\vartheta}\in\Lambda_{\vartheta}}\{(\boldsymbol{\lambda}_{\vartheta} - \hat{Z}^{j}_{T\vartheta\alpha})'(\hat{\mathcal{I}}^{-1}_{T\alpha})^{-1}_{\vartheta\vartheta}(\boldsymbol{\lambda}_{\vartheta} - \hat{Z}^{j}_{T\vartheta\alpha})\};$$

here the minimization of the quadratic form over the cone Λ_{ϑ} has to be performed numerically.

(iii) Set $\widehat{LR}^{j}_{T,A_G} = \max_{\alpha\in A_G}\widehat{LR}^{j}_{T}(\alpha)$.

This yields a sample $\{\widehat{LR}^{1}_{T,A_G}, \ldots, \widehat{LR}^{J}_{T,A_G}\}$ of J realizations. An approximate p–value corresponding to an observed LR test statistic LR_T is computed as $J^{-1}\sum_{j=1}^{J}\mathbf{1}(\widehat{LR}^{j}_{T,A_G} > LR_T)$ (here $\mathbf{1}(\cdot)$ denotes the indicator function). The precision of this approximation can be controlled by choosing J large enough, see Hansen (1996) (in the illustration below we use $J = 1000$).

4.2 A small Monte Carlo study

We now study the finite sample properties of the LR test statistics and the simulation-based critical values. The results are presented in Table 1. We consider two LR test statistics, one based on an estimated LMAR model, and another based on an estimated GMAR model (as in our two examples in the preceding sections). In all simulations, we use an autoregressive order $p = 1$, $J = 1000$ repetitions (see the previous subsection), and three different sample sizes: $T = 250$, 500, and 1000.

The top part of Table 1 presents results for size simulations. Data is generated from an AR(1) model (for a range of different parameter values shown in Table 1) and AR(1), LMAR(1), and GMAR(1) models are estimated (LMAR with $m = 1$; GMAR with the restriction $\tilde{\phi}_0 = \tilde{\varphi}_0$; in estimation of the mixture models we use a genetic algorithm as singularity of the information matrix may render gradient based methods unreliable). Two LR test statistics are calculated based on the estimated LMAR and GMAR models, respectively, and labelled 'LMAR LR_T' and 'GMAR LR_T'. Simulation-based p–values are computed based on the asymptotic distributions in Section 3.5.2 and using the simulation procedure in Section 4.1. Using nominal levels 10%, 5%, and 1%, a reject/not-reject decision is recorded. This exercise is repeated 1000 times, and the six rightmost columns in Table 1 present the empirical rejection frequencies (for the LMAR LR_T and GMAR LR_T tests and the three nominal levels used).

As can be seen from the results in Table 1 (top part), the LMAR LR_T test's size is satisfactory overall, typically being somewhat oversized for sample sizes $T = 250$ and 500, and somewhat conservative for the largest sample size ($T = 1000$). The parameter values used in simulation do not seem to have a large effect on the size. The GMAR LR_T test, on the other hand, appears to be moderately oversized across all sample sizes and parameter values used.

The lower part of Table 1 presents results for power simulations. Data is generated either from a GMAR model or from an LMAR model (for a range of different parameter values shown in Table 1), and empirical rejection frequencies are calculated as above. Both the LMAR LR_T test and the GMAR LR_T test appear to have good overall power. As expected, when the two regimes differ more from each other, the tests have higher power, and the same happens when sample size is increased. Besides having good power against the 'right' alternatives, the tests also turn out to have decent power against 'wrong' alternatives: When data is generated from the GMAR (resp., LMAR) model, the LMAR LR_T (resp., GMAR LR_T) test rejects reasonably often (the GMAR LR_T test in particular seems capable of picking up LMAR type regime switching). Naturally, the power of the tests may be inflated due to the tests being oversized.

As a computational remark we note that the LMAR LR_T and GMAR LR_T tests and their p–values are reasonably straightforward to compute in a matter of seconds using a standard, modern desktop computer (for one particular model, one particular sample size, and $J = 1000$ repetitions). The GMAR LR_T test is computationally more demanding than the LMAR LR_T test as it involves the minimization of a quadratic form over a cone which is not needed in the LMAR case (see Sections 3.5.2 and 4.1); this is also one potential reason for the less precise size of the GMAR LR_T test.

5 Conclusions

This paper has studied the asymptotic distribution of the LR test statistic for testing a linear autoregressive model against a two-regime mixture autoregressive model. A distinguishing feature of the paper is that the regime switching probabilities are observation-dependent. Technical challenges resulting from unidentified parameters under the null, parameters on the boundary, and singularity of the information matrix were dealt with by considering an appropriately reparameterized model and higher-order expansions of the log-likelihood function. The resulting asymptotic distribution of the LR test statistic is non-standard and application-specific. Critical values can be obtained by a straightforward simulation procedure, and a Monte Carlo study indicated the proposed tests to have satisfactory size and power properties.

The general theory of the paper was illustrated using two concrete examples, the LMAR model of Wong and Li (2001) and (a version of the) GMAR model of Kalliovirta et al. (2015). Considering other mixture AR models, as well as the general GMAR model, is left for future research. This paper was concerned with testing linearity against a two-regime model, and considering tests of $M \geq 2$ regimes versus $M + 1$ regimes, similarly as in Kasahara and Shimotsu (2015) in a related setting, forms another interesting research topic.

DGP	Parameter values								T	LMAR LR_T			GMAR LR_T		
										10%	5%	1%	10%	5%	1%
AR	$\tilde{\phi}_0$	$\tilde{\phi}_1$	$\tilde{\sigma}_1^2$												
	0	-0.75	1						250	0.13	0.07	0.014	0.18	0.11	0.028
									500	0.10	0.05	0.011	0.15	0.08	0.024
									1000	0.07	0.03	0.009	0.15	0.09	0.026
	0	-0.50	1						250	0.13	0.07	0.011	0.15	0.08	0.019
									500	0.10	0.06	0.012	0.15	0.08	0.024
									1000	0.08	0.04	0.008	0.14	0.09	0.020
	0	-0.25	1						250	0.14	0.07	0.022	0.16	0.09	0.020
									500	0.12	0.07	0.015	0.17	0.09	0.022
									1000	0.11	0.07	0.017	0.13	0.07	0.027
	0	0.00	1						250	0.12	0.06	0.009	0.14	0.07	0.023
									500	0.11	0.06	0.016	0.14	0.08	0.021
									1000	0.07	0.04	0.003	0.12	0.06	0.018
	0	0.25	1						250	0.14	0.09	0.022	0.16	0.09	0.026
									500	0.09	0.05	0.012	0.14	0.07	0.024
									1000	0.09	0.04	0.007	0.14	0.08	0.018
	0	0.50	1						250	0.14	0.08	0.023	0.17	0.10	0.031
									500	0.11	0.06	0.011	0.16	0.09	0.026
									1000	0.08	0.04	0.012	0.13	0.07	0.019
	0	0.75	1						250	0.12	0.06	0.012	0.16	0.10	0.024
									500	0.08	0.04	0.003	0.15	0.08	0.013
									1000	0.09	0.05	0.010	0.17	0.11	0.026
GMAR	$\tilde{\phi}_0$	$\tilde{\phi}_1$	$\tilde{\sigma}_1^2$	$\tilde{\varphi}_1$	$\tilde{\sigma}_2^2$	α									
	0	0.4	1	0.6	1	0.33			250	0.13	0.07	0.01	0.20	0.13	0.04
									500	0.11	0.07	0.01	0.19	0.12	0.04
									1000	0.10	0.06	0.01	0.27	0.18	0.06
	0	0.3	1	0.7	1	0.33			250	0.13	0.07	0.02	0.35	0.24	0.10
									500	0.17	0.10	0.03	0.50	0.36	0.17
									1000	0.18	0.12	0.04	0.69	0.58	0.32
	0	0.2	1	0.8	1	0.33			250	0.23	0.15	0.04	0.68	0.56	0.33
									500	0.33	0.23	0.09	0.88	0.81	0.60
									1000	0.49	0.40	0.20	0.99	0.98	0.92
	0	0.5	1	0.5	2	0.33			250	0.24	0.14	0.04	0.44	0.31	0.14
									500	0.26	0.17	0.05	0.55	0.42	0.20
									1000	0.32	0.21	0.09	0.70	0.60	0.34
	0	0.5	1	0.5	3	0.33			250	0.41	0.29	0.13	0.78	0.67	0.41
									500	0.57	0.46	0.25	0.92	0.86	0.70
									1000	0.74	0.67	0.46	0.99	0.98	0.94
LMAR	$\tilde{\phi}_0$	$\tilde{\phi}_1$	$\tilde{\sigma}_1^2$	$\tilde{\varphi}_0$	$\tilde{\varphi}_1$	$\tilde{\sigma}_2^2$	α_0	α_1							
	0	0.50	1	0	0.50	2	0	1	250	0.52	0.39	0.17	0.35	0.25	0.08
									500	0.74	0.64	0.40	0.45	0.35	0.14
									1000	0.94	0.90	0.76	0.61	0.47	0.23
	0	0.50	1	0	0.50	3	0	1	250	0.89	0.83	0.63	0.67	0.57	0.33
									500	0.99	0.98	0.93	0.88	0.80	0.61
									1000	1.00	1.00	1.00	0.98	0.96	0.89
	0	0.25	1	0	0.75	1	0	1	250	0.63	0.50	0.28	0.61	0.48	0.26
									500	0.87	0.81	0.64	0.80	0.70	0.47
									1000	1.00	0.99	0.96	0.95	0.91	0.76
	0	0.25	1	0	0.75	2	0	1	250	0.87	0.81	0.62	0.72	0.62	0.40
									500	0.99	0.98	0.95	0.91	0.84	0.64
									1000	1.00	1.00	1.00	0.99	0.98	0.92

Table 1: Results of a Monte Carlo study. Empirical rejection frequencies (six rightmost columns) for LMAR LR_T and GMAR LR_T tests for nominal sizes 10%, 5%, and 1%. Different rows correspond to results with data generated from AR, GMAR, or LMAR models (with parameter values used shown in the table) with different sample sizes ($T = 250$, 500, or 1000).

Appendix

A Details for the general results

Lemma A.1. *When Assumptions 2(ii) and 3(i,ii) hold, a sufficient condition for Assumption 3(iii) is that*

$$\|\boldsymbol{\pi}_\alpha(\phi,\varphi)-\boldsymbol{\pi}_\alpha(\phi^*,\phi^*)\|\leq Ch\left(\|(\phi,\varphi)-(\phi^*,\phi^*)\|_*\right)\ \text{ for all } (\phi,\varphi)\in\Phi\times\Phi,$$

where C is a finite positive constant, $h:[0,\infty)\to[0,\infty)$ is a strictly increasing function such that $h(x)\downarrow 0$ as $x\downarrow 0$, and $\|\cdot\|_*$ is any vector norm on $\mathbb{R}^{2q_2}$.

Proof of Lemma A.1. First note that $\sup_{\alpha\in A}\|(\hat\beta_{T\alpha},\hat\pi_{T\alpha},\hat\varpi_{T\alpha})-(\beta^*,\pi^*,0)\|\leq\sup_{\alpha\in A}\|\hat\beta_{T\alpha}-\beta^*\|+\sup_{\alpha\in A}\|(\hat\pi_{T\alpha},\hat\varpi_{T\alpha})-(\pi^*,0)\|$, where the former term on the majorant side is $o_p(1)$ by Assumption 2(ii). Second, the latter term equals $\sup_{\alpha\in A}\|\boldsymbol{\pi}_\alpha(\hat\phi_{T\alpha},\hat\varphi_{T\alpha})-\boldsymbol{\pi}_\alpha(\phi^*,\phi^*)\|$ which, due to the assumptions made in the lemma, can be bounded by $C\sup_{\alpha\in A}h\big(\|(\hat\phi_{T\alpha},\hat\varphi_{T\alpha})-(\phi^*,\phi^*)\|_*\big)\leq Ch\big(\sup_{\alpha\in A}\|(\hat\phi_{T\alpha},\hat\varphi_{T\alpha})-(\phi^*,\phi^*)\|_*\big)$. By Assumption 2(ii) and the fact that all vector norms on $\mathbb{R}^{2q_2}$ are equivalent, the majorant side is $o_p(1)$. ■

Proof of Lemma 1. Set $\boldsymbol{\theta}_{T\alpha}=\mathcal{I}_\alpha^{1/2}T^{1/2}\boldsymbol{\theta}(\alpha,\hat\beta_{T\alpha},\hat\pi_{T\alpha},\hat\varpi_{T\alpha})$, and rewrite equation (10) evaluated at $(\alpha,\hat\beta_{T\alpha},\hat\pi_{T\alpha},\hat\varpi_{T\alpha})$ as

$$o_{p\alpha}(1)\leq L_T^\pi(\alpha,\hat\beta_{T\alpha},\hat\pi_{T\alpha},\hat\varpi_{T\alpha})-L_T^\pi(\alpha,\beta^*,\pi^*,0)=(\mathcal{I}_\alpha^{1/2}Z_{T\alpha})'\boldsymbol{\theta}_{T\alpha}-\frac{1}{2}\|\boldsymbol{\theta}_{T\alpha}\|^2+R_T(\alpha,\hat\beta_{T\alpha},\hat\pi_{T\alpha},\hat\varpi_{T\alpha}) \tag{21}$$

(the lower bound is due to Assumption 2(i) and the definitions of $(\hat\beta_{T\alpha},\hat\pi_{T\alpha},\hat\varpi_{T\alpha})$ and $L_T^\pi(\alpha,\beta,\pi,\varpi)$). As to the first term on the right hand side, note that

$$\sup_{\alpha\in A}\|\mathcal{I}_\alpha^{1/2}Z_{T\alpha}\|=\sup_{\alpha\in A}\left(Z_{T\alpha}'\mathcal{I}_\alpha Z_{T\alpha}\right)^{1/2}\leq\sup_{\alpha\in A}\left(\lambda_{\max}\left(\mathcal{I}_\alpha\right)\right)^{1/2}\sup_{\alpha\in A}\|Z_{T\alpha}\|\leq C\sup_{\alpha\in A}\|Z_{T\alpha}\|=O_p\left(1\right), \tag{22}$$

where the latter inequality holds with some finite C in view of Assumption 5(iii), and the last equality will be justified below. Thus

$$\mathcal{I}_\alpha^{1/2}Z_{T\alpha}=O_{p\alpha}(1), \tag{23}$$

a result which also implies $\|(\mathcal{I}_\alpha^{1/2}Z_{T\alpha})'\boldsymbol{\theta}_{T\alpha}\|\leq\|\boldsymbol{\theta}_{T\alpha}\|\|\mathcal{I}_\alpha^{1/2}Z_{T\alpha}\|=\|\boldsymbol{\theta}_{T\alpha}\|O_{p\alpha}(1)$.

Next consider the third term on the right hand side of (21), where the assumption $(\hat\beta_{T\alpha},\hat\pi_{T\alpha},\hat\varpi_{T\alpha})=(\beta^*,\pi^*,0)+o_{p\alpha}(1)$ (see Assumption 3(iii)) allows us to choose a sequence $\{\gamma_T,\ T\geq 1\}$ of (non-random) positive scalars converging to zero slowly enough to ensure that $P(\sup_{\alpha\in A}\|(\hat\beta_{T\alpha},\hat\pi_{T\alpha},\hat\varpi_{T\alpha})-(\beta^*,\pi^*,0)\|\leq\gamma_T)\to 1$, and with this sequence $\{\gamma_T,\ T\geq 1\}$, Assumption 5(iv) implies that

$$|R_T(\alpha,\hat\beta_{T\alpha},\hat\pi_{T\alpha},\hat\varpi_{T\alpha})|=(1+\|T^{1/2}\boldsymbol{\theta}(\alpha,\hat\beta_{T\alpha},\hat\pi_{T\alpha},\hat\varpi_{T\alpha})\|)^2o_{p\alpha}(1) \tag{24}$$

(cf. Pakes and Pollard (1989, proof of Thm 3.3)). Here $\|T^{1/2}\boldsymbol{\theta}(\alpha,\hat\beta_{T\alpha},\hat\pi_{T\alpha},\hat\varpi_{T\alpha})\|=\|\mathcal{I}_\alpha^{-1/2}\boldsymbol{\theta}_{T\alpha}\|$ so that, as $0<\inf_{\alpha\in A}\lambda_{\min}(\mathcal{I}_\alpha)$ by Assumption 5(iii),

$$|R_T(\alpha,\hat\beta_{T\alpha},\hat\pi_{T\alpha},\hat\varpi_{T\alpha})|\leq o_{p\alpha}(1)+\|\boldsymbol{\theta}_{T\alpha}\|\,o_{p\alpha}(1)+\|\boldsymbol{\theta}_{T\alpha}\|^2\,o_{p\alpha}(1). \tag{25}$$

Combining the results above (see (21), the inequality below (23), and (25)), organizing terms, and absorbing $\|\boldsymbol{\theta}_{T\alpha}\| \, o_{p\alpha}(1)$ into $\|\boldsymbol{\theta}_{T\alpha}\| \, O_{p\alpha}(1)$ leads to

$$0 \leq \|\boldsymbol{\theta}_{T\alpha}\| \, O_{p\alpha}(1) + o_{p\alpha}(1) + (o_{p\alpha}(1) - \frac{1}{2}) \|\boldsymbol{\theta}_{T\alpha}\|^2 . \tag{26}$$

In Supplementary Appendix D we show that the last term on the right hand side of (26) is dominated by $-\frac{1}{4}\|\boldsymbol{\theta}_{T\alpha}\|^2 + o_{p\alpha}(1)$ so that (absorbing constants into the $O_{p\alpha}(1)$ and $o_{p\alpha}(1)$ terms) $\|\boldsymbol{\theta}_{T\alpha}\|^2 \leq 2\|\boldsymbol{\theta}_{T\alpha}\| \, O_{p\alpha}(1) + o_{p\alpha}(1)$. Denoting the $O_{p\alpha}(1)$ term on the majorant side with $\xi_{T\alpha}$ and reorganizing one obtains

$$(\|\boldsymbol{\theta}_{T\alpha}\| - \xi_{T\alpha})^2 = \|\boldsymbol{\theta}_{T\alpha}\|^2 - 2\|\boldsymbol{\theta}_{T\alpha}\| \, \xi_{T\alpha} + \xi_{T\alpha}^2 \leq \xi_{T\alpha}^2 + o_{p\alpha}(1) = O_{p\alpha}(1).$$

Taking square roots yields $\|\boldsymbol{\theta}_{T\alpha}\| = O_{p\alpha}(1)$ so that $\|T^{1/2}\boldsymbol{\theta}(\alpha, \hat{\beta}_{T\alpha}, \hat{\pi}_{T\alpha}, \hat{\varpi}_{T\alpha})\| = O_{p\alpha}(1)$ as claimed in assertion (i) of the lemma. Assertion (ii) of the lemma follows from the result $\|\boldsymbol{\theta}_{T\alpha}\| = O_{p\alpha}(1)$ and (25). Assertion (iii) follows directly from assertion (ii).

To complete the proof of Lemma 1, we now justify the last equality in (22). By Assumption 5(ii), $T^{-1/2}S_{T\bullet} \Rightarrow S_\bullet$ in $\mathcal{C}(A, \mathbb{R}^r)$, and by the continuous mapping theorem (justification in Supplementary Appendix D), $Z_{T\alpha} = \mathcal{I}_\alpha^{-1}T^{-1/2}S_{T\alpha}$ converges weakly in $\mathcal{C}(A, \mathbb{R}^r)$ to a mean zero $\mathbb{R}^r$-valued Gaussian process $Z_\alpha = \mathcal{I}_\alpha^{-1}S_\alpha$ whose sample paths are continuous in α with probability one and that has $E[Z_\alpha Z_\alpha'] = \mathcal{I}_\alpha^{-1}$ for all $\alpha \in A$. A further application of the continuous mapping theorem (justification in Supplementary Appendix D) implies that $\sup_{\alpha\in A}\|Z_{T\alpha}\|$ converges in distribution in $\mathbb{R}$ and, as all probability measures on $\mathbb{R}$ are tight, the limit must be tight. This justifies the last equality in (22). ■

Proof of Lemma 2. By the definition of $\hat{\boldsymbol{\lambda}}_{T\alpha q}$, the fact that $\mathbf{0} \in \Theta_{\alpha,T}$, and (23),

$$\left\|\mathcal{I}_\alpha^{1/2}(\hat{\boldsymbol{\lambda}}_{T\alpha q} - Z_{T\alpha})\right\|^2 = (\hat{\boldsymbol{\lambda}}_{T\alpha q} - Z_{T\alpha})'\mathcal{I}_\alpha(\hat{\boldsymbol{\lambda}}_{T\alpha q} - Z_{T\alpha}) \leq Z_{T\alpha}'\mathcal{I}_\alpha Z_{T\alpha} + o_{p\alpha}(1) = O_{p\alpha}(1),$$

implying that $\mathcal{I}_\alpha^{1/2}(\hat{\boldsymbol{\lambda}}_{T\alpha q} - Z_{T\alpha}) = O_{p\alpha}(1)$. Thus, a further use of (23) and the condition $0 < \inf_{\alpha\in A}\lambda_{\min}(\mathcal{I}_\alpha)$ in Assumption 5(iii) yields $\hat{\boldsymbol{\lambda}}_{T\alpha q} = O_{p\alpha}(1)$.

Next, we establish that $R_T(\alpha, \hat{\beta}_{T\alpha q}, \hat{\pi}_{T\alpha q}, \hat{\varpi}_{T\alpha q}) = o_{p\alpha}(1)$. First, $\hat{\boldsymbol{\lambda}}_{T\alpha q} = T^{1/2}\boldsymbol{\theta}(\alpha, \hat{\beta}_{T\alpha q}, \hat{\pi}_{T\alpha q}, \hat{\varpi}_{T\alpha q})$ and $\hat{\boldsymbol{\lambda}}_{T\alpha q} = O_{p\alpha}(1)$ imply that $\boldsymbol{\theta}(\alpha, \hat{\beta}_{T\alpha q}, \hat{\pi}_{T\alpha q}, \hat{\varpi}_{T\alpha q}) = o_{p\alpha}(1)$. Second, to show that $(\hat{\beta}_{T\alpha q}, \hat{\pi}_{T\alpha q}, \hat{\varpi}_{T\alpha q}) = (\beta^*, \pi^*, 0) + o_{p\alpha}(1)$, pick arbitrary $\epsilon, \delta > 0$, and conclude from Assumption 5(i)(b) that $\inf_{\alpha\in A}\inf_{(\beta,\pi,\varpi)\in B\times\Pi_\alpha:\|(\beta,\pi,\varpi)-(\beta^*,\pi^*,0)\|\geq\epsilon}\|\boldsymbol{\theta}(\alpha, \beta, \pi, \varpi)\| \geq \delta_\epsilon > 0$ for some δ_ϵ. Now, as $\boldsymbol{\theta}(\alpha, \hat{\beta}_{T\alpha q}, \hat{\pi}_{T\alpha q}, \hat{\varpi}_{T\alpha q}) = o_{p\alpha}(1)$, we can find a $T_{\delta,\delta_\epsilon}$ such that for all $T \geq T_{\delta,\delta_\epsilon}$, $P(\sup_{\alpha\in A}\|\boldsymbol{\theta}(\alpha, \hat{\beta}_{T\alpha q}, \hat{\pi}_{T\alpha q}, \hat{\varpi}_{T\alpha q})\| < \delta_\epsilon) > 1 - \delta$. Note that whenever the event $\{\sup_{\alpha\in A}\|\boldsymbol{\theta}(\alpha, \hat{\beta}_{T\alpha q}, \hat{\pi}_{T\alpha q}, \hat{\varpi}_{T\alpha q})\| < \delta_\epsilon\}$ occurs, the event $\{\sup_{\alpha\in A}\|(\hat{\beta}_{T\alpha q}, \hat{\pi}_{T\alpha q}, \hat{\varpi}_{T\alpha q}) - (\beta^*, \pi^*, 0)\| < \epsilon\}$ must also occur (if, on the contrary, $\|(\hat{\beta}_{T\alpha q}, \hat{\pi}_{T\alpha q}, \hat{\varpi}_{T\alpha q}) - (\beta^*, \pi^*, 0)\| \geq \epsilon$ for some $\alpha \in A$, then necessarily $\|\boldsymbol{\theta}(\alpha, \hat{\beta}_{T\alpha q}, \hat{\pi}_{T\alpha q}, \hat{\varpi}_{T\alpha q})\| \geq \delta_\epsilon$)). Therefore for all $T \geq T_{\delta,\delta_\epsilon}$, $1 - \delta < P(\sup_{\alpha\in A}\|\boldsymbol{\theta}(\alpha, \hat{\beta}_{T\alpha q}, \hat{\pi}_{T\alpha q}, \hat{\varpi}_{T\alpha q})\| < \delta_\epsilon) \leq P(\sup_{\alpha\in A}\|(\hat{\beta}_{T\alpha q}, \hat{\pi}_{T\alpha q}, \hat{\varpi}_{T\alpha q}) - (\beta^*, \pi^*, 0)\| < \epsilon)$, so that $\sup_{\alpha\in A}\|(\hat{\beta}_{T\alpha q}, \hat{\pi}_{T\alpha q}, \hat{\varpi}_{T\alpha q}) - (\beta^*, \pi^*, 0)\| = o_p(1)$, as desired.

Third, as $(\hat{\beta}_{T\alpha q}, \hat{\pi}_{T\alpha q}, \hat{\varpi}_{T\alpha q}) = (\beta^*, \pi^*, 0) + o_{p\alpha}(1)$, using the same argument as in the proof of Lemma 1 (see the derivation of equation (24)) now leads to

$$|R_T(\alpha, \hat{\beta}_{T\alpha q}, \hat{\pi}_{T\alpha q}, \hat{\varpi}_{T\alpha q})| = (1 + \|T^{1/2}\boldsymbol{\theta}(\alpha, \hat{\beta}_{T\alpha q}, \hat{\pi}_{T\alpha q}, \hat{\varpi}_{T\alpha q})\|)^2 o_{p\alpha}(1) = (1 + \|\hat{\boldsymbol{\lambda}}_{T\alpha q}\|)^2 o_{p\alpha}(1).$$

This, together with the result $\hat{\boldsymbol{\lambda}}_{T\alpha q} = O_{p\alpha}(1)$ established above, yields $R_T(\alpha, \hat{\beta}_{T\alpha q}, \hat{\pi}_{T\alpha q}, \hat{\varpi}_{T\alpha q}) = o_{p\alpha}(1)$.

Now, by expansion (11), the definitions of $(\hat{\beta}_{T\alpha}, \hat{\pi}_{T\alpha}, \hat{\varpi}_{T\alpha})$ and $\hat{\boldsymbol{\lambda}}_{T\alpha q}$, and making use of results

$R_T(\alpha, \hat{\beta}_{T\alpha}, \hat{\pi}_{T\alpha}, \hat{\varpi}_{T\alpha}) = o_{p\alpha}(1)$ and $R_T(\alpha, \hat{\beta}_{T\alpha q}, \hat{\pi}_{T\alpha q}, \hat{\varpi}_{T\alpha q}) = o_{p\alpha}(1)$,

$$\begin{aligned} o_{p\alpha}(1) &\leq L_T^{\pi}(\alpha, \hat{\beta}_{T\alpha}, \hat{\pi}_{T\alpha}, \hat{\varpi}_{T\alpha}) - L_T^{\pi}(\alpha, \hat{\beta}_{T\alpha q}, \hat{\pi}_{T\alpha q}, \hat{\varpi}_{T\alpha q}) \\ &= \frac{1}{2}(\hat{\boldsymbol{\lambda}}_{T\alpha q} - Z_{T\alpha})'\mathcal{I}_\alpha(\hat{\boldsymbol{\lambda}}_{T\alpha q} - Z_{T\alpha}) \\ &\quad - \frac{1}{2}\big[T^{1/2}\boldsymbol{\theta}(\alpha, \hat{\beta}_{T\alpha}, \hat{\pi}_{T\alpha}, \hat{\varpi}_{T\alpha}) - Z_{T\alpha}\big]'\mathcal{I}_\alpha\big[T^{1/2}\boldsymbol{\theta}(\alpha, \hat{\beta}_{T\alpha}, \hat{\pi}_{T\alpha}, \hat{\varpi}_{T\alpha}) - Z_{T\alpha}\big] + o_{p\alpha}(1) \\ &\leq o_{p\alpha}(1), \end{aligned}$$

implying, by (18), the desired result. ■

Proof of Lemma 3. For any vectors $a, b \in \mathbb{R}^r$, we denote $\|a-b\|_{\mathcal{I}_\alpha^{-1}} = [(a-b)'\mathcal{I}_\alpha(a-b)]^{1/2}$, and for any point $p \in \mathbb{R}^r$ and a set $S \subset \mathbb{R}^r$, we define $\|p - S\|_{\mathcal{I}_\alpha^{-1}}$ via $\|p-S\|^2_{\mathcal{I}_\alpha^{-1}} = \inf_{s\in S}\|s-p\|^2_{\mathcal{I}_\alpha^{-1}} = \inf_{s\in S}[(s-p)'\mathcal{I}_\alpha(s-p)]$. With this notation, we need to prove that $\|Z_{T\alpha} - \Theta_{\alpha,T}\|^2_{\mathcal{I}_\alpha^{-1}} = \|Z_{T\alpha} - \Lambda\|^2_{\mathcal{I}_\alpha^{-1}} + o_{p\alpha}(1)$. First note that, because Λ is a cone, we have, for any T,

$$\|Z_{T\alpha} - \Lambda\|^2_{\mathcal{I}_\alpha^{-1}} = T\inf_{\boldsymbol{\lambda}\in\Lambda}\big\{(T^{-1/2}\boldsymbol{\lambda} - T^{-1/2}Z_{T\alpha})'\mathcal{I}_\alpha(T^{-1/2}\boldsymbol{\lambda} - T^{-1/2}Z_{T\alpha})\big\} = T\big\|T^{-1/2}Z_{T\alpha} - \Lambda\big\|^2_{\mathcal{I}_\alpha^{-1}}.$$

Similarly, by the definitions of $\Theta_{\alpha,T}$ and Θ_α,

$$\|Z_{T\alpha} - \Theta_{\alpha,T}\|^2_{\mathcal{I}_\alpha^{-1}} = \inf_{\boldsymbol{\lambda}\in\Theta_\alpha}\big\{(T^{1/2}\boldsymbol{\lambda} - Z_{T\alpha})'\mathcal{I}_\alpha(T^{1/2}\boldsymbol{\lambda} - Z_{T\alpha})\big\} = T\big\|T^{-1/2}Z_{T\alpha} - \Theta_\alpha\big\|^2_{\mathcal{I}_\alpha^{-1}}.$$

Now let $G_T(\alpha, \boldsymbol{x}) = T\,\|\boldsymbol{x} - \Theta_\alpha\|^2_{\mathcal{I}_\alpha^{-1}} - T\,\|\boldsymbol{x} - \Lambda\|^2_{\mathcal{I}_\alpha^{-1}}$ define a (non-random) function on $A \times \mathbb{R}^r$. Because $\{\Theta_\alpha,\ \alpha \in A\}$ is locally uniformly equal to the cone Λ, we can find a $\delta > 0$ such that $\Theta_\alpha \cap (-\delta, \delta)^r = \Lambda \cap (-\delta,\delta)^r$ for all $\alpha \in A$. Furthermore, $\mathbf{0} \in \Theta_\alpha$ and $\mathbf{0} \in \Lambda$ (here $\mathbf{0} \in \mathbb{R}^r$). Therefore, we can find a neighborhood $N_{\mathbf{0}}$ of $\mathbf{0}$ such that for all $(\alpha, \boldsymbol{x}) \in A \times N_{\mathbf{0}}$,

$$G_T(\alpha, \boldsymbol{x}) = T\,\|\boldsymbol{x} - \Theta_\alpha \cap (-\delta,\delta)^r\|^2_{\mathcal{I}_\alpha^{-1}} - T\,\|\boldsymbol{x} - \Lambda \cap (-\delta,\delta)^r\|^2_{\mathcal{I}_\alpha^{-1}} = 0.$$

Now define $G_T(\alpha)$, a random function of α, as $G_T(\alpha) = G_T(\alpha, T^{-1/2}Z_{T\alpha})$. In the proof of Lemma 1 it was shown that $\sup_{\alpha\in A}\|Z_{T\alpha}\| = O_p(1)$ (see (22)) so that $T^{-1/2}\|Z_{T\alpha}\| = o_{p\alpha}(1)$. Therefore, for all $\epsilon > 0$,

$$\begin{aligned} P\big(\sup_{\alpha\in A}|G_T(\alpha)| > \epsilon\big) &\leq P\big(\sup_{\alpha\in A}|G_T(\alpha)| > \epsilon\ ;\ T^{-1/2}\sup_{\alpha\in A}\|Z_{T\alpha}\| \in N_0\big) \\ &\quad + P\big(\sup_{\alpha\in A}|G_T(\alpha)| > \epsilon\ ;\ T^{-1/2}\sup_{\alpha\in A}\|Z_{T\alpha}\| \notin N_0\big) \\ &= P\big(\sup_{\alpha\in A}|G_T(\alpha)| > \epsilon\ ;\ T^{-1/2}\sup_{\alpha\in A}\|Z_{T\alpha}\| \notin N_0\big) \\ &\leq P\big(T^{-1/2}\sup_{\alpha\in A}\|Z_{T\alpha}\| \notin N_0\big) \\ &\to 0, \end{aligned}$$

where the equality holds because $G_T(\alpha, \boldsymbol{x}) = 0$ for all $(\alpha, \boldsymbol{x}) \in A \times N_{\mathbf{0}}$, and the convergence holds because $T^{-1/2}\|Z_{T\alpha}\| = o_{p\alpha}(1)$. Thus $\sup_{\alpha\in A}|G_T(\alpha)| = o_p(1)$, implying the desired result $\|Z_{T\alpha} - \Theta_{\alpha,T}\|^2_{\mathcal{I}_\alpha^{-1}} = \|Z_{T\alpha} - \Lambda\|^2_{\mathcal{I}_\alpha^{-1}} + o_{p\alpha}(1)$. ■

Proof of Lemma 4. It was shown in the proof of Lemma 1 that $Z_{T\bullet} \Rightarrow Z_\bullet$ in $\mathcal{C}(A, \mathbb{R}^r)$. Therefore also $(Z_{T\bullet}, \mathcal{I}_\bullet) \Rightarrow (Z_\bullet, \mathcal{I}_\bullet)$ in $\mathcal{C}(A,\mathbb{R}^r) \times \{\mathcal{I}_\alpha\}$ (Billingsley (1999, Thm. 3.9)). As the function $g : \mathcal{C}(A, \mathbb{R}^r) \times \{\mathcal{I}_\alpha\} \to \mathcal{B}(A, \mathbb{R})$ mapping $(x_\bullet, \mathcal{I}_\bullet)$ ($\in \mathcal{C}(A,\mathbb{R}^r) \times \{\mathcal{I}_\alpha\}$) to $x_\bullet'\mathcal{I}_\bullet x_\bullet - \inf_{\boldsymbol{\lambda}\in\Lambda}\{(\boldsymbol{\lambda} - x_\bullet)'\mathcal{I}_\bullet(\boldsymbol{\lambda} - x_\bullet)\}$ is continuous (justification in Supplementary Appendix D), the continuous mapping theorem is applicable. This, together with Billingsley (1999, Thm 3.1) (for which it is necessary that the remainder term in (20) is $o_{p\alpha}(1)$ and not only $o_p(1)$), implies that

$$2[L_T^{\pi}(\bullet,\hat{\beta}_{T\bullet},\hat{\pi}_{T\bullet},\hat{\varpi}_{T\bullet}) - L_T^{\pi}(\bullet,\beta^*,\pi^*,0)] \Rightarrow Z_\bullet'\mathcal{I}_\bullet Z_\bullet - \inf_{\boldsymbol{\lambda}\in\Lambda}\left\{(\boldsymbol{\lambda}-Z_\bullet)'\mathcal{I}_\bullet(\boldsymbol{\lambda}-Z_\bullet)\right\},$$

establishing the desired result. ■

Proof of Lemma 5. The proof consists of reasonably straightforward matrix algebra. For details, see Supplementary Appendix D. ■

Proof of Lemma 6. The required arguments are standard but presented for completeness and to contrast them with arguments that lead to Lemma 4. The reparameterization described in Assumption 3 is unnecessary and the original $\tilde{\phi}$–parameterization may be used (alternatively, consider the identity mapping $\pi = \boldsymbol{\pi}(\tilde{\phi}) = \tilde{\phi}$). As for the quadratic expansion of the log-likelihood function, let $\boldsymbol{\theta}(\tilde{\phi}) = (\tilde{\phi} - \tilde{\phi}^*)$ take the role of $\boldsymbol{\theta}(\alpha,\beta,\pi,\varpi)$, and note that straightforward derivations (similar to those used in the LMAR example in Section 3.3.1) yield

$$\begin{aligned} L_T^0(\tilde{\phi}) - L_T^0(\tilde{\phi}^*) &= (T^{-1/2}S_T^0)'[T^{1/2}\boldsymbol{\theta}(\tilde{\phi})] - \frac{1}{2}[T^{1/2}\boldsymbol{\theta}(\tilde{\phi})]'\mathcal{I}^0[T^{1/2}\boldsymbol{\theta}(\tilde{\phi})] + R_T(\tilde{\phi}), \\ R_T(\tilde{\phi}) &= \frac{1}{2}[T^{1/2}\boldsymbol{\theta}(\tilde{\phi})]'[T^{-1}\nabla_{\phi\phi'}L_T^0(\dot{\phi}) - (-\mathcal{I}^0)][T^{1/2}\boldsymbol{\theta}(\tilde{\phi})], \end{aligned}$$

with $\dot{\phi}$ denoting a point between $\tilde{\phi}$ and $\tilde{\phi}^*$. Validity of Assumption 5 follows from the arguments used in connection with the LMAR example together with Assumption 8. Assumption 6 holds with $\Lambda = \mathbb{R}^{p+2}$. Arguments analogous to those that lead to Lemma 4 now yield the stated convergence result, and the convergence is joint as in both cases it follows from the weak convergence result $T^{-1/2}S_{T\bullet} \Rightarrow S_\bullet$. ■

Proof of Theorem 1. Under Assumption 8, the random process $S_{\theta\alpha}'\mathcal{I}_{\theta\theta\alpha}^{-1}S_{\theta\alpha}$ in Lemma 5 coincides with the random variable $S^{0\prime}(\mathcal{I}^0)^{-1}S^0$ in Lemma 6. Therefore the expression of $LR_T(\alpha)$ in (7), Lemmas 4, 5, and 6, and Billingsley (1999, Thm 3.1) (for which it is necessary that the remainder term in (7) is $o_{p\alpha}(1)$ and not only $o_p(1)$) imply the weak convergence result for $LR_T(\alpha)$. The result for LR_T follows from the continuous mapping theorem. ■

B Details for the LMAR example

In this appendix it appears convenient to denote $\alpha_{1,t}^L$ instead of α_t^L and to set $\alpha_{2,t}^L = 1 - \alpha_{1,t}^L$. In some cases we also include the argument α and denote $\alpha_{1,t}^L(\alpha)$ and $\alpha_{2,t}^L(\alpha)$. The same notation is employed in the Supplementary Appendix and a similar modification is used in the case of the GMAR model.

Assumptions 1–4. Assumption 1(i) is assumed to hold, 1(ii) holds as A is compact, and 1(iii) holds by the definition of the mixing weight. For the verification of Assumption 2, see the GMAR example in Appendix C; the LMAR case is treated there as well. To verify Assumption 3, note first that conditions (i) and (ii) clearly hold, and for condition (iii), we have

$$\boldsymbol{\pi}_\alpha(\phi,\varphi) - \boldsymbol{\pi}_\alpha(\phi^*,\phi^*) = (\phi,\phi-\varphi) - (\phi^*,0) = (\phi-\phi^*,(\phi-\phi^*)-(\varphi-\phi^*))\,\alpha_{1,t}^L\,.$$

Choosing $\|x\|_* = \|x\|_1 = \sum_{i=1}^{2q} |x_i|$ and using the triangle inequality it is straightforward to check that

$$\|(\phi - \phi^*, (\phi - \phi^*) - (\varphi - \phi^*))\| \leq 2\|\phi - \phi^*\| + \|\varphi - \phi^*\| \leq 2\|(\phi, \varphi) - (\phi^*, \phi^*)\|_1 .$$

Thus, Assumption 3(iii) holds by Lemma A.1. Regarding Assumption 4, as $\alpha_{1,t}^L$ does not depend on (ϕ, φ) and $\boldsymbol{\pi}_\alpha^{-1}(\pi, \varpi) = (\pi, \pi - \varpi)$, the required differentiability conditions hold for all positive integers k.

Assumption 5: Computation of the required derivatives. As $\alpha_{1,t}^L$ does not depend on (ϕ, φ) and $\boldsymbol{\pi}_\alpha^{-1}(\pi, \varpi) = (\pi, \pi - \varpi)$, the quantities $f_{2,t}^\pi(\alpha, \pi, \varpi)$ and $l_t^\pi(\alpha, \pi, \varpi)$ take the form

$$\begin{aligned} f_{2,t}^\pi(\alpha, \pi, \varpi) &= \alpha_{1,t}^L(\alpha) f_t(\pi) + (1 - \alpha_{1,t}^L(\alpha)) f_t(\pi - \varpi), \\ l_t^\pi(\alpha, \pi, \varpi) &= \log[\alpha_{1,t}^L(\alpha) f_t(\pi) + (1 - \alpha_{1,t}^L(\alpha)) f_t(\pi - \varpi)]. \end{aligned}$$

Straightforward differentiation yields the following expressions for the first and second partial derivatives with respect to π and ϖ (recall that ∇ and ∇^2 denote first and second order differentiation with respect to the indicated parameters, and $\nabla f_t(\cdot)$ denotes differentiation of $f_t(\cdot)$ in (3) with respect to $\tilde{\phi} = (\tilde{\phi}_0, \tilde{\phi}_1, \ldots, \tilde{\phi}_p, \tilde{\sigma}_1^2)$):

$$\begin{aligned} \nabla_\pi l_t^\pi(\alpha, \pi, \varpi) &= [\alpha_{1,t}^L(\alpha) \nabla f_t(\pi) + \alpha_{2,t}^L(\alpha) \nabla f_t(\pi - \varpi)] / f_{2,t}^\pi(\alpha, \pi, \varpi), \\ \nabla_\varpi l_t^\pi(\alpha, \pi, \varpi) &= [-\alpha_{2,t}^L(\alpha) \nabla f_t(\pi - \varpi)] / f_{2,t}^\pi(\alpha, \pi, \varpi), \\ \nabla^2_{\pi\pi'} l_t^\pi(\alpha, \pi, \varpi) &= \alpha_{1,t}^L(\alpha) \left(\frac{\nabla^2 f_t(\pi)}{f_{2,t}^\pi(\alpha, \pi, \varpi)} - \frac{\nabla f_t(\pi)}{f_{2,t}^\pi(\alpha, \pi, \varpi)} \frac{\alpha_{1,t}^L(\alpha) \nabla' f_t(\pi) + \alpha_{2,t}^L(\alpha) \nabla' f_t(\pi - \varpi)}{f_{2,t}^\pi(\alpha, \pi, \varpi)} \right) \\ &\quad + \alpha_{2,t}^L(\alpha) \left(\frac{\nabla^2 f_t(\pi - \varpi)}{f_{2,t}^\pi(\alpha, \pi, \varpi)} - \frac{\nabla f_t(\pi - \varpi)}{f_{2,t}^\pi(\alpha, \pi, \varpi)} \frac{\alpha_{1,t}^L(\alpha) \nabla' f_t(\pi) + \alpha_{2,t}^L(\alpha) \nabla' f_t(\pi - \varpi)}{f_{2,t}^\pi(\alpha, \pi, \varpi)} \right), \\ \nabla^2_{\pi\varpi'} l_t^\pi(\alpha, \pi, \varpi) &= \alpha_{1,t}^L(\alpha) \left(-\frac{\nabla f_t(\pi)}{f_{2,t}^\pi(\alpha, \pi, \varpi)} \frac{-\alpha_{2,t}^L(\alpha) \nabla' f_t(\pi - \varpi)}{f_{2,t}^\pi(\alpha, \pi, \varpi)} \right) \\ &\quad + \alpha_{2,t}^L(\alpha) \left(-\frac{\nabla^2 f_t(\pi - \varpi)}{f_{2,t}^\pi(\alpha, \pi, \varpi)} + \frac{\nabla f_t(\pi - \varpi)}{f_{2,t}^\pi(\alpha, \pi, \varpi)} \frac{\alpha_{2,t}^L(\alpha) \nabla' f_t(\pi - \varpi)}{f_{2,t}^\pi(\alpha, \pi, \varpi)} \right), \\ \nabla^2_{\varpi\varpi'} l_t^\pi(\alpha, \pi, \varpi) &= -\alpha_{2,t}^L(\alpha) \left(-\frac{\nabla^2 f_t(\pi - \varpi)}{f_{2,t}^\pi(\alpha, \pi, \varpi)} + \frac{\nabla f_t(\pi - \varpi)}{f_{2,t}^\pi(\alpha, \pi, \varpi)} \frac{\alpha_{2,t}^L(\alpha) \nabla' f_t(\pi - \varpi)}{f_{2,t}^\pi(\alpha, \pi, \varpi)} \right). \end{aligned}$$

The corresponding expressions evaluated at $(\alpha, \pi, \varpi) = (\alpha, \pi^*, 0)$ take the form

$$\nabla_{(\pi,\varpi)} l_t^\pi(\alpha, \pi^*, 0) = (1, -(1 - \alpha_{1,t}^L(\alpha))) \otimes \frac{\nabla f_t(\pi^*)}{f_t(\pi^*)}, \tag{27}$$

$$\nabla^2_{\pi\pi'} l_t^\pi(\alpha, \pi^*, 0) = \frac{\nabla^2 f_t(\pi^*)}{f_t(\pi^*)} - \frac{\nabla f_t(\pi^*)}{f_t(\pi^*)} \frac{\nabla' f_t(\pi^*)}{f_t(\pi^*)}, \tag{28}$$

$$\nabla^2_{\pi\varpi'} l_t^\pi(\alpha, \pi^*, 0) = -(1 - \alpha_{1,t}^L(\alpha)) \frac{\nabla^2 f_t(\pi^*)}{f_t(\pi^*)} + (1 - \alpha_{1,t}^L(\alpha)) \frac{\nabla f_t(\pi^*)}{f_t(\pi^*)} \frac{\nabla' f_t(\pi^*)}{f_t(\pi^*)}, \tag{29}$$

$$\nabla^2_{\varpi\varpi'} l_t^\pi(\alpha, \pi^*, 0) = (1 - \alpha_{1,t}^L(\alpha)) \frac{\nabla^2 f_t(\pi^*)}{f_t(\pi^*)} - (1 - \alpha_{1,t}^L(\alpha))^2 \frac{\nabla f_t(\pi^*)}{f_t(\pi^*)} \frac{\nabla' f_t(\pi^*)}{f_t(\pi^*)}. \tag{30}$$

Assumption 5: Verifying the assumption. Omitting the unnecessary β from $\boldsymbol{\theta}(\alpha,\beta,\pi,\varpi)$, we have $\boldsymbol{\theta}(\alpha,\pi,\varpi)=(\pi-\pi^*,\varpi)=(\theta,\vartheta)$ so that part (i) is clearly satisfied with the parameter space

$$\begin{aligned}\Theta_\alpha=\Theta&=\{\boldsymbol{\theta}=(\theta,\vartheta)\in\mathbb{R}^{2q_2}:\theta=\pi-\pi^*,\vartheta=\varpi\text{ for some }(\pi,\varpi)\in\Pi\}\\&=\{\boldsymbol{\theta}=(\theta,\vartheta)\in\mathbb{R}^{2q_2}:\theta=\phi-\phi^*,\vartheta=\phi-\varphi\text{ for some }(\phi,\varphi)\in\Phi\times\Phi\}\end{aligned}$$

independent of α and with 0 ($\in\mathbb{R}^{2q_2}$) an interior point of Θ. The first two requirements in part (ii) are similarly clear, whereas the third requirement follows from the continuity of $\alpha^L_{1,t}(\alpha)$ in α. The weak convergence requirement in part (ii) is verified in Supplementary Appendix E.1.

Now consider part (iii) of Assumption 5, and first consider the positive definiteness of $\mathcal{I}_\alpha$ for a fixed $\alpha\in A$. It suffices to show that

$$a'\frac{\nabla f_t(\pi^*)}{f_t(\pi^*)}=\alpha^L_{2,t}(\alpha)b'\frac{\nabla f_t(\pi^*)}{f_t(\pi^*)}\ \text{ a.s.}\tag{31}$$

only if $a=(a_1,\ldots,a_{q_2})=0$ and $b=(b_1,\ldots,b_{q_2})=0$. For brevity, denote $g_t(\pi)=[y_t-(\pi_1+\pi_2y_{t-1}+\cdots+\pi_{p+1}y_{t-p})]/\pi_{p+2}^{1/2}=[y_t-(\tilde{\phi}_0+\tilde{\phi}_1y_{t-1}+\cdots+\tilde{\phi}_py_{t-p})]/\tilde{\sigma}_1$ and $\boldsymbol{z}_{t-1}=(1,\boldsymbol{y}_{t-1})$ so that $g_t(\pi^*)=\varepsilon_t$. Straightforward differentiation yields

$$\frac{\nabla f_t(\pi)}{f_t(\pi)}=\begin{bmatrix}\frac{1}{\tilde{\sigma}_1}\boldsymbol{z}_{t-1}g_t(\pi)\\\frac{1}{2\tilde{\sigma}_1^2}(g_t^2(\pi)-1)\end{bmatrix}\text{ so that }\frac{\nabla f_t(\pi^*)}{f_t(\pi^*)}=\begin{bmatrix}\frac{1}{\tilde{\sigma}_1^*}\boldsymbol{z}_{t-1}\varepsilon_t\\\frac{1}{2\tilde{\sigma}_1^{*2}}(\varepsilon_t^2-1)\end{bmatrix}\tag{32}$$

(where $\tilde{\sigma}_1^2=\pi_{p+2}$). Multiplying both sides of equation (31) by $(\varepsilon_t^2-1)2\tilde{\sigma}_1^{*2}$, taking expectations conditional on $\mathcal{F}_{t-1}$, and making use of the fact that odd moments of the normal distribution are zero, yields $a_{q_2}E[(\varepsilon_t^2-1)^2]=\alpha^L_{2,t}(\alpha)b_{q_2}E[(\varepsilon_t^2-1)^2]$ a.s. Because $\alpha^L_{2,t}(\alpha)\neq0$ and not equal to a constant (see Section 3.1.1), it follows that $a_{q_2}=b_{q_2}=0$. Therefore, equation (31) (multiplied by σ_1^*) now reduces to $(a_1,\ldots,a_{q_2-1})'\boldsymbol{z}_{t-1}\varepsilon_t=\alpha^L_{2,t}(\alpha)(b_1,\ldots,b_{q_2-1})'\boldsymbol{z}_{t-1}\varepsilon_t$ a.s. Multiplying this equation by ε_t, taking expectations conditional on $\mathcal{F}_{t-1}$, and dividing by $E[\varepsilon_t^2]=\sigma_1^{*2}$ yields $(a_1,\ldots,a_{q_2-1})'\boldsymbol{z}_{t-1}=\alpha^L_{2,t}(\alpha)(b_1,\ldots,b_{q_2-1})'\boldsymbol{z}_{t-1}$ a.s. This is clearly impossible unless $a_1=\ldots=a_{q_2-1}=b_1=\ldots=b_{q_2-1}=0$, because $\alpha^L_{2,t}(\alpha)$ is a positive and strictly decreasing function of $\alpha'\boldsymbol{z}_{t-1}$ ($\neq\alpha_0$) and because $(a_1,\ldots,a_{q_2-1})'\boldsymbol{z}_{t-1}$ and $(b_1,\ldots,b_{q_2-1})'\boldsymbol{z}_{t-1}$ are normally distributed or constants (if only a_1 and b_1 are nonzero). Therefore, $a=b=0$, so that $\mathcal{I}_\alpha$ is positive definite (for any fixed $\alpha\in A$).

To complete the verification of part (iii), we show that $\mathcal{I}_\alpha$ is a continuous function of α and such that $0<\inf_{\alpha\in A}\lambda_{\min}(\mathcal{I}_\alpha)$ and $\sup_{\alpha\in A}\lambda_{\max}(\mathcal{I}_\alpha)<\infty$. For continuity, let α_n be a sequence of points in A converging to $\alpha_\bullet\in A$. It suffices to demonstrate that $\lim_{n\to\infty}E\big[\alpha^L_{2,t}(\alpha_n)\frac{\nabla f_t(\pi^*)}{f_t(\pi^*)}\frac{\nabla' f_t(\pi^*)}{f_t(\pi^*)}\big]=E\big[\alpha^L_{2,t}(\alpha_\bullet)\frac{\nabla f_t(\pi^*)}{f_t(\pi^*)}\frac{\nabla' f_t(\pi^*)}{f_t(\pi^*)}\big]$ and similarly with $\alpha^L_{2,t}(\cdot)$ replaced by its square. This, however, is an immediate consequence of the dominated convergence theorem because $\alpha^L_{2,t}(\alpha)$ is a continuous positive function of α and smaller than 1, and because $E\big[\|\frac{\nabla f_t(\pi^*)}{f_t(\pi^*)}\frac{\nabla' f_t(\pi^*)}{f_t(\pi^*)}\|\big]<\infty$ due to Lemma F.1 (in Supplementary Appendix F.5). The statements on the eigenvalues follow from the continuity of $\mathcal{I}_\alpha$, the compactness of its domain A, and the positive definiteness of $\mathcal{I}_\alpha$ for all fixed $\alpha\in A$ shown above.

As for part (iv) of Assumption 5, based on the expression of the remainder term in (15) it suffices to show that for all sequences of (non-random) positive scalars $\{\gamma_T,\ T\geq1\}$ for which $\gamma_T\to0$ as $T\to\infty$,

$$\sup_{(\pi,\varpi)\in\Pi:\|(\pi,\varpi)-(\pi^*,0)\|\leq\gamma_T}\big\|T^{-1}\nabla^2_{(\pi,\varpi)(\pi,\varpi)'}L^\pi_T(\alpha,\pi,\varpi)-(-\mathcal{I}_\alpha)\big\|=o_{p\alpha}(1),\tag{33}$$

where the parameter space of (π,ϖ), denoted by Π, is independent of α, as noted above. First we show that a uniform law of large numbers applies to the matrix $T^{-1}\nabla^2_{(\pi,\varpi)(\pi,\varpi)'}L^\pi_T(\alpha,\pi,\varpi)$ on $A\times\Pi$, that is,

$$\sup_{\alpha\in A}\sup_{(\pi,\varpi)\in\Pi}\left\|T^{-1}\nabla^2_{(\pi,\varpi)(\pi,\varpi)'}L^{\pi}_T(\alpha,\pi,\varpi)-E[\nabla^2_{(\pi,\varpi)(\pi,\varpi)'}l^{\pi}_t(\alpha,\pi,\varpi)]\right\|=o_p(1). \tag{34}$$

As $T^{-1}\nabla^2_{(\pi,\varpi)(\pi,\varpi)'}L^{\pi}_T(\alpha,\pi,\varpi)=T^{-1}\sum_{t=1}^T\nabla^2_{(\pi,\varpi)(\pi,\varpi)'}l^{\pi}_t(\alpha,\pi,\varpi)$ with $\nabla^2_{(\pi,\varpi)(\pi,\varpi)'}l^{\pi}_t(\alpha,\pi,\varpi)$ a function of the stationary and ergodic process $(y_t,\boldsymbol{y}_{t-1})$ (by Assumption 1(i)), we only need to establish that $E\big[\sup_{\alpha\in A}\sup_{(\pi,\varpi)\in\Pi}\|\nabla^2_{(\pi,\varpi)(\pi,\varpi)'}l^{\pi}_t(\alpha,\pi,\varpi)\|\big]<\infty$ (see Ranga Rao (1962)). Verification of this moment condition is provided in Supplementary Appendix E.2. Furthermore, using the dominated convergence theorem and arguments similar to those used in part (iii), it can be shown that $E[\nabla^2_{(\pi,\varpi)(\pi,\varpi)'}l^{\pi}_t(\alpha,\pi,\varpi)]$ is a (uniformly) continuous function of (α,π,ϖ) on $A\times\Pi$. Now note that the left hand side of (33) is dominated by

$$\begin{aligned}&\sup_{(\pi,\varpi)\in\Pi}\left\|T^{-1}\nabla^2_{(\pi,\varpi)(\pi,\varpi)'}L^{\pi}_T(\alpha,\pi,\varpi)-E[\nabla^2_{(\pi,\varpi)(\pi,\varpi)'}l^{\pi}_t(\alpha,\pi,\varpi)]\right\|\\&\quad+\sup_{(\pi,\varpi)\in\Pi:\|(\pi,\varpi)-(\pi^*,0)\|\le\gamma_T}\left\|E[\nabla^2_{(\pi,\varpi)(\pi,\varpi)'}l^{\pi}_t(\alpha,\pi,\varpi)]-(-\mathcal{I}_\alpha)\right\|.\end{aligned}$$

The former term is, due to (34), of order $o_{p\alpha}(1)$. Regarding the latter term, the supremum of it over $\alpha\in A$ converges to zero due to the uniform continuity of $E[\nabla^2_{(\pi,\varpi)(\pi,\varpi)'}l^{\pi}_t(\alpha,\pi,\varpi)]$ and the fact that $E[\nabla^2_{(\pi,\varpi)(\pi,\varpi)'}l^{\pi}_t(\alpha,\pi^*,0)]=-\mathcal{I}_\alpha$ (this fact follows from the expression of $\nabla^2_{(\pi,\varpi)(\pi,\varpi)'}l^{\pi}_t(\alpha,\pi^*,0)$ given in (28)–(30) and Lemma F.3). Thus, we have verified (33), and hence Assumption 5(iv).

Assumptions 6–8. That Θ is locally (uniformly) equal to the cone $\Lambda=\mathbb{R}^{2q}$ follows from the expression of the set Θ given in the verification of Assumption 5 above and the fact that 0 ($\in\mathbb{R}^{2q}$) is an interior point of Θ. Assumption 7 is clear, as Assumption 6 holds with the cone $\Lambda=\mathbb{R}^{2q}$. Assumption 8 is clear from the verification of Assumption 5.

Expression of $\hat{s}_{t\alpha}$ in Section 4.1. Let $\hat{\varepsilon}_t$ denote the OLS residuals rescaled by the estimated standard deviation, i.e., $\hat{\varepsilon}_t=(y_t-\hat{\hat{\phi}}_{T,0}-\hat{\hat{\phi}}_{T,1}y_{t-1}-\cdots-\hat{\hat{\phi}}_{T,p}y_{t-p})/\hat{\hat{\sigma}}_T$, and set $\hat{s}_{t\alpha}=\big(\nabla f_t(\hat{\hat{\phi}}_T)/f_t(\hat{\hat{\phi}}_T),-(1-\alpha^L_{1,t}(\alpha))\nabla f_t(\hat{\hat{\phi}}_T)/f_t(\hat{\hat{\phi}}_T)\big)$ with $\nabla f_t(\hat{\hat{\phi}}_T)/f_t(\hat{\hat{\phi}}_T)=(\frac{1}{\hat{\hat{\sigma}}_T}\boldsymbol{z}_{t-1}\hat{\varepsilon}_t,\frac{1}{2\hat{\hat{\sigma}}^2_T}(\hat{\varepsilon}^2_t-1))$ (see (13) and (32)).

C Details for the GMAR example

Assumption 1. Assumption 1(i) is assumed to hold. Assumption 1(ii) holds as A is a compact subset of $(0,1)$. Assumption 1(iii) holds by the definition of the mixing weight.

Assumption 2. For each fixed $\alpha\in A$, compactness of B and Φ together with the continuity of $L_T(\alpha,\beta,\phi,\varphi)=\sum_{t=1}^T l_t(\alpha,\beta,\phi,\varphi)$ ensures the existence of a measurable maximizer $(\hat{\beta}_{T\alpha},\hat{\phi}_{T\alpha},\hat{\varphi}_{T\alpha})$. Hence part (i) holds (with the $o_{p\alpha}(1)$ term equal to zero). (In the GMAR case, this maximizer is not unique when $\alpha=1/2$, but this does not matter for Assumption 2.)

To prove that $\sup_{\alpha\in A}||(\hat{\beta}_{T\alpha},\hat{\phi}_{T\alpha},\hat{\varphi}_{T\alpha})-(\beta^*,\phi^*,\phi^*)||\xrightarrow{p}0$, by Andrews (1993, Lemma A1) it suffices to show that (a) $\sup_{(\alpha,\beta,\phi,\varphi)\in A\times B\times\Phi\times\Phi}\left|T^{-1}L_T(\alpha,\beta,\phi,\varphi)-E[l_t(\alpha,\beta,\phi,\varphi)]\right|\xrightarrow{p}0$ as $T\to\infty$ and that (b) for every neighborhood $N(\beta^*,\phi^*,\phi^*)$ of (β^*,ϕ^*,ϕ^*),

$$\sup_{\alpha\in A}\sup_{(\beta,\phi,\varphi)\in B\times\Phi\times\Phi\setminus N(\beta^*,\phi^*,\phi^*)}(E[l_t(\alpha,\beta,\phi,\varphi)]-E[l_t(\alpha,\beta^*,\phi^*,\phi^*)])<0.$$

Property (a) can be verified by using the uniform law of large numbers given in Ranga Rao (1962). As $T^{-1}L_T(\alpha,\beta,\phi,\varphi)=T^{-1}\sum_{t=1}^T l_t(\alpha,\beta,\phi,\varphi)$ with $l_t(\alpha,\beta,\phi,\varphi)$ a function of the stationary and ergodic process $(y_t,\boldsymbol{y}_{t-1})$, we only need to show that $E\big[\sup_{(\alpha,\beta,\phi,\varphi)\in A\times B\times\Phi\times\Phi}|l_t(\alpha,\beta,\phi,\varphi)|\big]<\infty$.

Making use of Assumption 1, it is easy to show that $C_1 \exp\{-C_2(1+y_t^2+\cdots+y_{t-p}^2)\}) \leq f_t(\beta,\phi) \leq C_2$ for some $0 < C_1, C_2 < \infty$ and for all $(\beta,\phi) \in B \times \Phi$, so that $\log(C_1) - C_2(1+y_t^2+\cdots+y_{t-p}^2) \leq l_t(\alpha,\beta,\phi,\varphi) \leq \log(C_2)$ for all $(\alpha,\beta,\phi,\varphi) \in A \times B \times \Phi \times \Phi$ (cf. Kalliovirta et al. (2015, pp. 264–265)); this holds in both the LMAR and GMAR cases. The required moment condition follows from this.

As for property (b), the uniform law of large numbers used above also delivers the continuity of the limit function $E[l_t(\alpha,\beta,\phi,\varphi)]$ on the compact set $A \times B \times \Phi \times \Phi$. Therefore it suffices to show that, for each fixed $\alpha \in A$, $E[l_t(\alpha,\beta,\phi,\varphi)] - E[l_t(\alpha,\beta^*,\phi^*,\phi^*)] \leq 0$ with equality if and only if $(\beta,\phi,\varphi) = (\beta^*,\phi^*,\phi^*)$. For the GMAR model, this can be straightforwardly shown with arguments used in the proof of Theorem 2 in Kalliovirta et al. (2015). To this end, define

$$\mathsf{n}_1(\nu_{1,t} \mid \boldsymbol{\nu}_{1,t-1};(\beta,\phi)) = (2\bar{\pi}\tilde{\sigma}_1^2)^{-1/2} \exp\left(-\frac{(\nu_{1,t} - \tilde{\phi}_0 - \sum_{i=1}^{p} \tilde{\phi}_i \nu_{1,t-i})^2}{2\tilde{\sigma}_1^2}\right),$$

where $\nu_{1,t}$ is the auxiliary Gaussian AR(p) process introduced in the GMAR example of Section 2.2. Clearly, $\mathsf{n}_1(\nu_{1,t} \mid \boldsymbol{\nu}_{1,t-1};(\beta,\phi))$ is the conditional density of $\nu_{1,t}$ given $\boldsymbol{\nu}_{1,t-1} = (\nu_{1,t-1},\ldots,\nu_{1,t-p})$. The notation $\mathsf{n}_1(\nu_{2,t} \mid \boldsymbol{\nu}_{2,t-1};(\beta,\varphi))$ is defined similarly by using the parameters $\tilde{\varphi}_i$ and $\tilde{\sigma}_2^2$ instead of $\tilde{\phi}_i$ and $\tilde{\sigma}_1^2$.

Now, in the same way as in the above-mentioned proof of Kalliovirta et al. (2015) we can use arguments based on the Kullback-Leibler divergence and conclude that, for each fixed $\alpha \in A$, $E[l_t(\alpha,\beta,\phi,\varphi)] - E[l_t(\alpha,\beta^*,\phi^*,\phi^*)] \leq 0$ with equality if and only if for almost all $(y,\boldsymbol{y}) \in \mathbb{R} \times \mathbb{R}^p$

$$\alpha_1^G \mathsf{n}_1(y \mid \boldsymbol{y};(\beta,\phi)) + \alpha_2^G \mathsf{n}_1(y \mid \boldsymbol{y};(\beta,\varphi)) = \mathsf{n}_1(y \mid \boldsymbol{y};(\beta^*,\phi^*)), \tag{35}$$

where we use α_m^G to stand for $\alpha_{m,t}^G$ but with $\boldsymbol{y}_{t-1}$ therein replaced by $\boldsymbol{y}$ ($m = 1,2$). Using well-known results on identification of finite mixtures of Gaussian distributions we find that, for each fixed $\alpha \in A$, and for each fixed $\boldsymbol{y} \in \mathbb{R}^p$ at a time, $\mathsf{n}_1(y \mid \boldsymbol{y};(\beta,\phi)) = \mathsf{n}_1(y \mid \boldsymbol{y};(\beta,\varphi)) = \mathsf{n}_1(y \mid \boldsymbol{y};(\beta^*,\phi^*))$ for almost all y. Using the arguments following equation (A.4) in Kalliovirta et al. (2015) we can now establish the desired result $(\beta,\phi,\varphi) = (\beta^*,\phi^*,\phi^*)$.

The arguments used for the GMAR model above can also be used for the LMAR model, but two things are worth noting. First, the proof given for the GMAR model above goes through even when there are no common parameters so that ϕ and φ could be used in place of (β,ϕ) and (β,φ). Second, equation (35) can be obtained in the same way as in the GMAR case even though the derivation of the related equation (A.4) in Kalliovirta et al. (2015) made use of the explicit expression of the stationary density of $(y_t,\boldsymbol{y}_{t-1})$ which is known for the GMAR model but, in general, unknown for the LMAR model. The reason for this is that the null hypothesis is here assumed to hold so that y_t is a linear Gaussian AR(p) process, implying that $(y_t,\boldsymbol{y}_{t-1})$ is normally distributed with density function a $p+1$ dimensional counterpart of the p dimensional normal density function $\mathsf{n}_p(\boldsymbol{\nu}_{1,t};\tilde{\phi})$ defined in equation (6) (see the GMAR example of Section 2.2). After observing these two facts we can proceed in the same way as in the GMAR case and conclude that equation (35) holds also for the LMAR model as long as we replace the mixing weights of the GMAR model with those of the LMAR model. As the arguments employed in the proof of the GMAR case after equation (35) made no use of the mixing weights they apply also to the LMAR model and can be used to complete the proof.

Assumptions 3 and 4. Conditions (i) and (ii) of Assumption 3 clearly hold and condition (iii) can be verified in the same way as in the case of the LMAR model. Specifically, we have

$$\boldsymbol{\pi}_\alpha(\phi,\varphi) - \boldsymbol{\pi}_\alpha(\phi^*,\phi^*) = (\alpha\phi + (1-\alpha)\varphi, \phi-\varphi) - (\phi^*,0) = (\alpha(\phi-\phi^*) + (1-\alpha)(\varphi-\phi^*), (\phi-\phi^*) - (\varphi-\phi^*)),$$

and choosing $\|x\|_* = \|x\|_1 = \sum_{i=1}^{2q} |x_i|$ it can straightforwardly be seen that condition (iii) holds by Lemma A.1. Regarding Assumption 4, based on the expression of α_t^G and the definition of $\boldsymbol{\pi}_\alpha^{-1}(\pi,\varpi) = (\pi + (1-\alpha)\varpi, \pi - \alpha\varpi)$, the required differentiability holds for all positive integers k.

Assumption 5: Derivation of expansion (10). First note that, as $\boldsymbol{\pi}_\alpha^{-1}(\pi,\varpi) = (\pi+(1-\alpha)\varpi, \pi-\alpha\varpi)$, the reparameterized mixing weight in the GMAR model is given by

$$\alpha_{1,t}^{G\pi}(\alpha,\beta,\pi,\varpi) = \frac{\alpha \mathsf{n}_p(\boldsymbol{y}_{t-1};(\beta,\pi+(1-\alpha)\varpi))}{\alpha \mathsf{n}_p(\boldsymbol{y}_{t-1};(\beta,\pi+(1-\alpha)\varpi)) + (1-\alpha)\mathsf{n}_p(\boldsymbol{y}_{t-1};(\beta,\pi-\alpha\varpi))}$$

and the quantities $f_{2,t}^\pi(\alpha,\beta,\pi,\varpi)$ and $l_t^\pi(\alpha,\beta,\pi,\varpi)$ take the form

$$\begin{aligned} f_{2,t}^\pi(\alpha,\beta,\pi,\varpi) &= \alpha_{1,t}^{G\pi}(\alpha,\beta,\pi,\varpi) f_t(\beta,\pi+(1-\alpha)\varpi) + (1-\alpha_{1,t}^{G\pi}(\alpha,\beta,\pi,\varpi)) f_t(\beta,\pi-\alpha\varpi), \\ l_t^\pi(\alpha,\beta,\pi,\varpi) &= \log[\alpha_{1,t}^{G\pi}(\alpha,\beta,\pi,\varpi) f_t(\beta,\pi+(1-\alpha)\varpi) + (1-\alpha_{1,t}^{G\pi}(\alpha,\beta,\pi,\varpi)) f_t(\beta,\pi-\alpha\varpi)]. \end{aligned}$$

Partial derivatives of the (reparameterized) log-likelihood function (with respect to (β,π,ϖ)) can be obtained with straightforward differentation but, as the calculations are somewhat lengthy, they are relegated to Supplementary Appendix F.1.

Now consider, for an arbitrary fixed $\alpha \in A$, a standard fourth-order Taylor expansion of $L_T^\pi(\alpha,\beta,\pi,\varpi) = \sum_{t=1}^T l_t^\pi(\alpha,\beta,\pi,\varpi)$ around $(\beta^*,\pi^*,0)$ with respect to the parameters (β,π,ϖ). For brevity, we write $\tilde{\pi} = (\beta,\pi)$ and $\tilde{\pi}^* = (\beta^*,\pi^*)$. Collecting terms that turn out to be asymptotically negligible into a remainder term yields

$$\begin{aligned} &L_T^\pi(\alpha,\tilde{\pi},\varpi) - L_T^\pi(\alpha,\tilde{\pi}^*,0) \\ &= (\tilde{\pi}-\tilde{\pi}^*)'\nabla_{\tilde{\pi}} L_T^\pi(\alpha,\tilde{\pi}^*,0) + \frac{1}{2!}(\tilde{\pi}-\tilde{\pi}^*)'\nabla^2_{\tilde{\pi}\tilde{\pi}'} L_T^\pi(\alpha,\tilde{\pi}^*,0)(\tilde{\pi}-\tilde{\pi}^*) \\ &+ \frac{1}{2!}\varpi'\nabla^2_{\varpi\varpi'} L_T^\pi(\alpha,\tilde{\pi}^*,0)\varpi + \frac{3}{3!}\sum_{i=1}^{q_1+q_2}\sum_{j=1}^{q_2}\sum_{k=1}^{q_2} \nabla^3_{\tilde{\pi}_i\varpi_j\varpi_k} L_T^\pi(\alpha,\tilde{\pi}^*,0)(\tilde{\pi}_i-\tilde{\pi}_i^*)\varpi_j\varpi_k \\ &+ \frac{1}{4!}\sum_{i=1}^{q_2}\sum_{j=1}^{q_2}\sum_{k=1}^{q_2}\sum_{l=1}^{q_2} \nabla^4_{\varpi_i\varpi_j\varpi_k\varpi_l} L_T^\pi(\alpha,\tilde{\pi}^*,0)\varpi_i\varpi_j\varpi_k\varpi_l + R_T^{(1)}(\alpha,\tilde{\pi},\varpi) \end{aligned} \quad (36)$$

with an explicit expression of the remainder term $R_T^{(1)}(\alpha,\tilde{\pi},\varpi)$ available in Supplementary Appendix F.2. Therein we also demonstrate that this fourth-order Taylor expansion can be written as a quadratic expansion of the form (10) given by

$$\begin{aligned} &L_T^\pi(\alpha,\beta,\pi,\varpi) - L_T^\pi(\alpha,\beta^*,\pi^*,0) \\ &= S_T'\boldsymbol{\theta}(\alpha,\beta,\pi,\varpi) - \frac{1}{2}[T^{1/2}\boldsymbol{\theta}(\alpha,\beta,\pi,\varpi)]'\mathcal{I}[T^{1/2}\boldsymbol{\theta}(\alpha,\beta,\pi,\varpi)] + R_T(\alpha,\beta,\pi,\varpi) \end{aligned} \quad (37)$$

or, setting $Z_T = \mathcal{I}^{-1}T^{-1/2}S_T$, in an alternative form corresponding to (11), with an explicit expression of the remainder term $R_T(\alpha,\beta,\pi,\varpi)$ available in Supplementary Appendix F.2.

The required derivatives are available in Supplementary Appendix F.1. Here we only present the derivatives that appear in the expression of S_T in (16), that is, the components of $\tilde{\nabla}_{\boldsymbol{\theta}} l_t^{\pi*} = (\tilde{\nabla}_\theta l_t^{\pi*}, \tilde{\nabla}_\vartheta l_t^{\pi*})$. From Supplementary Appendix F.1 we obtain (here ∇_i denotes the ith component of a derivative, and $\nabla \mathsf{n}_p(\cdot)$ denotes differentiation of $\mathsf{n}_p(\cdot)$ in (6) with respect to $\tilde{\phi} = (\tilde{\phi}_0, \tilde{\phi}_1, \ldots, \tilde{\phi}_p, \tilde{\sigma}_1^2))$

$$\nabla_\beta l_t^\pi(\alpha, \beta^*, \pi^*, 0) = \frac{\nabla_1 f_t^*}{f_t^*} \tag{38}$$

$$\nabla_{\pi_i} l_t^\pi(\alpha, \beta^*, \pi^*, 0) = \frac{\nabla_{i+1} f_t^*}{f_t^*} \tag{39}$$

$$\nabla^2_{\varpi_i \varpi_j} l_t^\pi(\alpha, \beta^*, \pi^*, 0) = \alpha(1-\alpha)\left[\frac{\nabla_{i+1}\mathsf{n}_p^*}{\mathsf{n}_p^*}\frac{\nabla_{j+1} f_t^*}{f_t^*} + \frac{\nabla_{i+1} f_t^*}{f_t^*}\frac{\nabla_{j+1}\mathsf{n}_p^*}{\mathsf{n}_p^*} + \frac{\nabla^2_{i+1,j+1} f_t^*}{f_t^*}\right] \tag{40}$$

where $i, j = 1, \ldots, p+1$, and, for brevity, we denote $f_t^* = f_t(\beta^*, \pi^*)$, $\mathsf{n}_p^* = \mathsf{n}_p(\beta^*, \pi^*)$, and similarly for their derivatives. Explicit expressions for the derivatives of f_t and n_p are given in Supplementary Appendix F.3.

Assumption 5: Verifying the assumption. As $\boldsymbol{\theta}(\alpha, \beta, \pi, \varpi) = (\beta - \beta^*, \pi - \pi^*, \alpha(1-\alpha)v(\varpi))$, part (i.a) is clearly satisfied. Assumption 1 requires α to be bounded away from zero and one, so that also part (i.b) is satisfied.

For part (ii), notice from (16), (17), and (38)–(40) that $S_T = \sum_{t=1}^T s_t$ with

$$s_t = \left(\frac{\nabla f_t^*}{f_t^*}, c_{11} X^*_{t,1,1}, \ldots, c_{q_2-1,q_2} X^*_{t,q_2-1,q_2}\right) \tag{41}$$

where, for $i, j \in \{1, \ldots, q_2\}$, the c_{ij}'s are as in Section 3.3.1 ($c_{ij} = 1/2$ if $i = j$ and $c_{ij} = 1$ if $i \neq j$) and

$$X^*_{t,i,j} = \frac{\nabla_{i+1}\mathsf{n}_p^*}{\mathsf{n}_p^*}\frac{\nabla_{j+1} f_t^*}{f_t^*} + \frac{\nabla_{i+1} f_t^*}{f_t^*}\frac{\nabla_{j+1}\mathsf{n}_p^*}{\mathsf{n}_p^*} + \frac{\nabla^2_{i+1,j+1} f_t^*}{f_t^*} \tag{42}$$

(note that s_t, and hence S_T, does not involve α as it cancels out from the expressions in (17) and (40)). Therefore the first three requirements in part (ii) are clearly satisfied. For the weak convergence requirement in part (ii) it now suffices to show that $T^{-1/2} S_T \xrightarrow{d} S$ in $\mathbb{R}^r$ for some multivariate Gaussian random vector S with mean zero and $E[SS'] = \mathcal{I}$. To this end, s_t clearly forms a stationary and ergodic process. Moreover, due to Lemma F.3 in Supplementary Appendix F.5, $E[\nabla_i f_t^*/f_t^* \mid \boldsymbol{y}_{t-1}] = E[\nabla^2_{ij} f_t^*/f_t^* \mid \boldsymbol{y}_{t-1}] = 0$ for any $i, j \in \{1, \ldots, q_2\}$ so that s_t is a martingale difference sequence. From the expression of $\mathcal{I}$ in (49) in Supplementary Appendix F.2 it is clear that $E[s_t s_t'] = \mathcal{I}$. Positive definiteness of $\mathcal{I}$ is proven in Supplementary Appendix F.3. The stated convergence result now follows from the central limit theorem of Billingsley (1961) in conjunction with the Cramér-Wold device.

For part (iii), it suffices to show the finiteness and positive definiteness of $\mathcal{I}$; these are proven in Supplementary Appendices F.2 and F.3. Part (iv) is proven in Supplementary Appendix F.4.

Assumption 6. By the definition of the set Θ_α and the transformation $(\phi, \varphi) \to (\pi, \varpi)$, the set Θ_α (see Sec 3.3.1) can equivalently be expressed as

$$\Theta_\alpha = \{\boldsymbol{\theta} = (\theta, \vartheta) \in \mathbb{R}^{q_1+q_2+q_\vartheta} : \theta = (\beta - \beta^*, \alpha(\phi - \phi^*) + (1-\alpha)(\varphi - \phi^*)), \vartheta = \alpha(1-\alpha)v(\phi - \varphi)$$
$$\text{for some } (\beta, \phi, \varphi) \in B \times \Phi^2\}.$$

We aim to show that the collection of sets $\{\Theta_\alpha,\ \alpha \in A\}$ is locally uniformly equal to the cone $\Lambda = \mathbb{R}^{q_1+q_2} \times v(\mathbb{R}^{q_2})$ where $v(\mathbb{R}^{q_2}) = \{v(\varpi) : \varpi \in \mathbb{R}^{q_2}\}$.

Let $\bar{S}((\beta^*, \phi^*), \delta)$ denote a closed $(q_1 + q_2)$-sphere centered at (β^*, ϕ^*) and with radius δ, and $\bar{S}(\phi^*, \delta)$ a similar q_2-sphere. As (β^*, ϕ^*) is an interior point of $B \times \Phi$, we can find a $\delta_1 > 0$ such that $\bar{S}((\beta^*, \phi^*), \delta_1) \times \bar{S}(\phi^*, \delta_1) \subset B \times \Phi^2$. By the definitions of the transformations $(\alpha, \beta, \phi, \varphi) \to \theta(\alpha, \beta, \phi, \varphi)$

and $(\alpha, \beta, \phi, \varphi) \to \vartheta(\alpha, \beta, \phi, \varphi)$ (defined implicitly by the definition of Θ_α above), we can find a $\delta_2 > 0$ such that

$$(-\delta_2, \delta_2)^{q_1+q_2} \times v((-\delta_2, \delta_2)^{q_2}) \subset \bigcap_{\alpha \in A} \Theta_\alpha(\delta_1) \subset \bigcap_{\alpha \in A} \Theta_\alpha$$

where

$$\begin{aligned}\Theta_\alpha(\delta_1) = \{\boldsymbol{\theta} = (\theta, \vartheta) \in \mathbb{R}^{q_1+q_2+q_\vartheta} : \theta = (\beta - \beta^*, \alpha(\phi - \phi^*) + (1-\alpha)(\varphi - \phi^*)), \vartheta = \alpha(1-\alpha)v(\phi - \varphi) \\ \text{for some } (\beta, \phi, \varphi) \in \bar{S}((\beta^*, \phi^*), \delta_1) \times \bar{S}(\phi^*, \delta_1)\}.\end{aligned}$$

Thus $(-\delta_2, \delta_2)^{q_1+q_2} \times v((-\delta_2, \delta_2)^{q_2}) \subset \Theta_\alpha$ for all $\alpha \in A$ so that (assuming, without loss of generality, that $\delta_2 < 1$)

$$(-\delta_2^2, \delta_2^2)^{q_1+q_2+q_\vartheta} \cap \Theta_\alpha = (-\delta_2^2, \delta_2^2)^{q_1+q_2} \times v((-\delta_2, \delta_2)^{q_2}) = (-\delta_2^2, \delta_2^2)^{q_1+q_2+q_\vartheta} \cap [\mathbb{R}^{q_1+q_2} \times v(\mathbb{R}^{q_2})]$$

for all $\alpha \in A$. Thus the collection of sets $\{\Theta_\alpha,\ \alpha \in A\}$ is locally uniformly equal to the cone $\Lambda = \mathbb{R}^{q_1+q_2} \times v(\mathbb{R}^{q_2})$.

Assumptions 7 and 8. Assumption 7 is satisfied as Assumption 6 holds with the cone $\Lambda = \mathbb{R}^{q_1+q_2} \times v(\mathbb{R}^{q_2})$ (note that now $q_\vartheta = q_2(q_2+1)/2$). Assumption 8 is clear from the verification of Assumption 5.

Expression of $\hat{s}_{t\alpha}$ in Section 4.1. Let $\hat{\varepsilon}_t$ and $\nabla f_t(\hat{\hat{\phi}}_T)/f_t(\hat{\hat{\phi}}_T)$ be as in the LMAR example (see Appendix B) and set (see (41) and (42)) $\hat{s}_t = \big(\nabla f_t(\hat{\hat{\phi}}_T)/f_t(\hat{\hat{\phi}}_T), c_{11}X_{t,1,1}(\hat{\hat{\phi}}_T), \ldots, c_{q_2-1,q_2}X_{t,q_2-1,q_2}(\hat{\hat{\phi}}_T)\big)$ where, for $i, j \in \{1, \ldots, q_2\}$, the c_{ij}'s are as in Section 3.3.1 ($c_{ij} = 1/2$ if $i = j$ and $c_{ij} = 1$ if $i \neq j$) and

$$X_{t,i,j}(\hat{\hat{\phi}}_T) = \frac{\nabla_{i+1}\mathsf{n}_p(\hat{\hat{\phi}}_T)}{\mathsf{n}_p(\hat{\hat{\phi}}_T)} \frac{\nabla_{j+1}f_t(\hat{\hat{\phi}}_T)}{f_t(\hat{\hat{\phi}}_T)} + \frac{\nabla_{i+1}f_t(\hat{\hat{\phi}}_T)}{f_t(\hat{\hat{\phi}}_T)} \frac{\nabla_{j+1}\mathsf{n}_p(\hat{\hat{\phi}}_T)}{\mathsf{n}_p(\hat{\hat{\phi}}_T)} + \frac{\nabla^2_{i+1,j+1}f_t(\hat{\hat{\phi}}_T)}{f_t(\hat{\hat{\phi}}_T)}.$$

Explicit expressions for the elements of $\nabla^2 f_t(\hat{\hat{\phi}}_T)/f_t(\hat{\hat{\phi}}_T)$ can be obtained from (50) in Supplementary Appendix F.3 by replacing ε_t and $\tilde{\sigma}_1^*$ therein with $\hat{\varepsilon}_t$ and $\hat{\hat{\sigma}}_T$, respectively. Expressions for the elements of $\nabla \mathsf{n}_p(\hat{\hat{\phi}}_T)/\mathsf{n}_p(\hat{\hat{\phi}}_T)$ can be obtained by evaluating (51) in Supplementary Appendix F.3 at $\tilde{\phi} = \hat{\hat{\phi}}_T$.

References

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

Andrews, D. W., 1993. Tests for parameter instability and structural change with unknown change point. Econometrica 61, 821–856.

Andrews, D. W., 1999. Estimation when a parameter is on a boundary. Econometrica 67, 1341–1383.

Andrews, D. W., 2000. Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space. Econometrica 68, 399–405.

Andrews, D. W., 2001. Testing when a parameter is on the boundary of the maintained hypothesis. Econometrica 69, 683–734.

Andrews, D. W., Ploberger, W., 1995. Admissibility of the likelihood ratio test when a nuisance parameter is present only under the alternative. Annals of Statistics 23, 1609–1629.

Bec, F., Rahbek, A., Shephard, N., 2008. The ACR model: a multivariate dynamic mixture autoregression. Oxford Bulletin of Economics and Statistics 70, 583–618.

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

Billingsley, P., 1999. Convergence of Probability Measures, 2nd Edition. Wiley.

Carrasco, M., Hu, L., Ploberger, W., 2014. Optimal test for Markov switching parameters. Econometrica 82, 765–784.

Cho, J. S., White, H., 2007. Testing for regime switching. Econometrica 75, 1671–1720.

Davies, R. B., 1977. Hypothesis testing when a nuisance parameter is present only under the alternative. Biometrika 64, 247–254.

Davies, R. B., 1987. Hypothesis testing when a nuisance parameter is present only under the alternative. Biometrika 74, 33–43.

Diebold, F. X., Lee, J.-H., Weinbach, G. C., 1994. Regime switching with time-varying transition probabilities. In: Hargreaves, C. (Ed.), Nonstationary Time Series Analysis and Cointegration. Oxford University Press, pp. 283–302.

Dueker, M. J., Psaradakis, Z., Sola, M., Spagnolo, F., 2011. Multivariate contemporaneous-threshold autoregressive models. Journal of Econometrics 160, 311–325.

Dueker, M. J., Sola, M., Spagnolo, F., 2007. Contemporaneous threshold autoregressive models: estimation, testing and forecasting. Journal of Econometrics 141, 517–547.

Filardo, A. J., 1994. Business-cycle phases and their transitional dynamics. Journal of Business & Economic Statistics 12, 299–308.

Frühwirth-Schnatter, S., 2006. Finite Mixture and Markov Switching Models. Springer.

Garcia, R., 1998. Asymptotic null distribution of the likelihood ratio test in Markov switching models. International Economic Review 39, 763–788.

Gouriéroux, C., Robert, C. Y., 2006. Stochastic unit root models. Econometric Theory 22, 1052–1090.

Hallin, M., Ley, C., 2014. Skew-symmetric distributions and Fisher information: the double sin of the skew-normal. Bernoulli 20, 1432–1453.

Hamilton, J. D., 2016. Macroeconomic regimes and regime shifts. In: Uhlig, H., Taylor, J. (Eds.), Handbook of Macroeconomics. Vol. 2A. Elsevier, pp. 163–201.

Hansen, B. E., 1992. The likelihood ratio test under nonstandard conditions: testing the Markov switching model of GNP. Journal of Applied Econometrics 7, S61–S82.

Hansen, B. E., 1996. Inference when a nuisance parameter is not identified under the null hypothesis. Econometrica 64, 413–430.

Jeffries, N. O., 1998. Logistic mixtures of generalized linear model times series. Ph.D. thesis, University of Maryland, College Park.

Kalliovirta, L., Meitz, M., Saikkonen, P., 2015. A Gaussian mixture autoregressive model for univariate time series. Journal of Time Series Analysis 36, 247–266.

Kalliovirta, L., Meitz, M., Saikkonen, P., 2016. Gaussian mixture vector autoregression. Journal of Econometrics 192, 485–498.

Kasahara, H., Shimotsu, K., 2012. Testing the number of components in finite mixture models, unpublished working paper.

Kasahara, H., Shimotsu, K., 2015. Testing the number of components in normal mixture regression models. Journal of the American Statistical Association 110, 1632–1645.

Kasahara, H., Shimotsu, K., 2017. Testing the number of regimes in Markov regime switching models, unpublished working paper.

Kim, C.-J., Piger, J., Startz, R., 2008. Estimation of Markov regime-switching regression models with endogenous switching. Journal of Econometrics 143, 263–273.

Lanne, M., Saikkonen, P., 2003. Modeling the US short-term interest rate by mixture autoregressive processes. Journal of Financial Econometrics 1, 96–125.

Lütkepohl, H., 2005. New Introduction to Multiple Time Series Analysis. Springer.

McLachlan, G., Peel, D., 2000. Finite Mixture Models. Wiley.

Pakes, A., Pollard, D., 1989. Simulation and the asymptotics of optimization estimators. Econometrica 57, 1027–1057.

Qu, Z., Zhuo, F., 2017. Likelihood ratio based tests for Markov regime switching, unpublished working paper.

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

Rotnitzky, A., Cox, D. R., Bottai, M., Robins, J., 2000. Likelihood-based inference with singular information matrix. Bernoulli 6, 243–284.

Shen, J., He, X., 2015. Inference for subgroup analysis with a structured logistic-normal mixture model. Journal of the American Statistical Association 110, 303–312.

Wong, C. S., Li, W. K., 2000. On a mixture autoregressive model. Journal of the Royal Statistical Society: Series B 62, 95–115.

Wong, C. S., Li, W. K., 2001. On a logistic mixture autoregressive model. Biometrika 88, 833–846.

Zhu, H., Zhang, H., 2004. Hypothesis testing in mixture regression models. Journal of the Royal Statistical Society: Series B 66, 3–16.

Zhu, H., Zhang, H., 2006. Asymptotics for estimation and testing procedures under loss of identifiability. Journal of Multivariate Analysis 97, 19–45.

Supplementary Appendix to 'Testing for observation-dependent regime switching in mixture autoregressive models' by Meitz and Saikkonen (not meant for publication)

D Further details for the general results

Proof of Lemma 1, further details. To justify that the last term on the right hand side of (26) is dominated by $-\frac{1}{4}\left\|\boldsymbol{\theta}_{T\alpha}\right\|^2 + o_{p\alpha}(1)$, note first that

$$\left(o_{p\alpha}(1) - \frac{1}{2}\right)\left\|\boldsymbol{\theta}_{T\alpha}\right\|^2 = \left(o_{p\alpha}(1) - \frac{1}{4}\right)\left\|\boldsymbol{\theta}_{T\alpha}\right\|^2 - \frac{1}{4}\left\|\boldsymbol{\theta}_{T\alpha}\right\|^2 := W_{T\alpha}\left\|\boldsymbol{\theta}_{T\alpha}\right\|^2 - \frac{1}{4}\left\|\boldsymbol{\theta}_{T\alpha}\right\|^2,$$

where $W_{T\alpha} = -\frac{1}{4} + o_{p\alpha}(1)$. Thus, $P\left(\sup_{\alpha\in A} W_{T\alpha} \leq 0\right) \to 1$ and (here $\mathbf{1}(\cdot)$ denotes the indicator function)

$$\begin{aligned}\sup_{\alpha\in A} W_{T\alpha}\left\|\boldsymbol{\theta}_{T\alpha}\right\|^2 &= \sup_{\alpha\in A} W_{T\alpha}\left\|\boldsymbol{\theta}_{T\alpha}\right\|^2 \mathbf{1}\left(\sup_{\alpha\in A} W_{T\alpha} \leq 0\right) + \sup_{\alpha\in A} W_{T\alpha}\left\|\boldsymbol{\theta}_{T\alpha}\right\|^2 \mathbf{1}\left(\sup_{\alpha\in A} W_{T\alpha} > 0\right) \\ &\leq \sup_{\alpha\in A} W_{T\alpha}\left\|\boldsymbol{\theta}_{T\alpha}\right\|^2 \mathbf{1}\left(\sup_{\alpha\in A} W_{T\alpha} > 0\right),\end{aligned}$$

where the last term is non-negative and positive with probability that is at most $P\left(\sup_{\alpha\in A} W_{T\alpha} > 0\right) \to 0$. Thus, combining the above derivations yields the desired result $\left(o_{p\alpha}(1) - \frac{1}{2}\right)\left\|\boldsymbol{\theta}_{T\alpha}\right\|^2 \leq -\frac{1}{4}\left\|\boldsymbol{\theta}_{T\alpha}\right\|^2 + o_{p\alpha}(1)$.

To justify the use of the continuous mapping theorem, note that in the first instance it is applied with the function $g : \mathcal{C}(A,\mathbb{R}^r)\times\{\mathcal{I}_\alpha\} \to \mathcal{C}(A,\mathbb{R}^r)$ mapping $(x_\bullet, \mathcal{I}_\bullet)$ to $\mathcal{I}_\bullet^{-1}x_\bullet$. Here $\mathcal{I}_\alpha^{-1}x_\alpha$ is continuous in α by Assumption 5(iii). Also, the latter set in the product $\mathcal{C}(A,\mathbb{R}^r)\times\{\mathcal{I}_\alpha\}$ contains only the non-random function $\mathcal{I}_\alpha$; this product space can be equipped with essentially the same metric as $\mathcal{C}(A,\mathbb{R}^r)$; cf. Andrews and Ploberger (1994, p. 1392 and 1407) and Zhu and Zhang (2006, proof of Theorem 5). In the second instance, the continuous mapping theorem is applied with the function $g : \mathcal{B}(A,\mathbb{R}^r) \to \mathbb{R}$ mapping $x_\bullet$ ($\in \mathcal{B}(A,\mathbb{R}^r)$) to $\sup_{\alpha\in A}\|x_\alpha\|$. For continuity, we need to establish that if a sequence $x_{n\bullet}$ converges to $x_\bullet$ in $\mathcal{B}(A,\mathbb{R}^r)$, then $g(x_{n\bullet}) \to g(x_\bullet)$ in $\mathbb{R}$ (i.e., if $\sup_{\alpha\in A}\|x_{n\alpha} - x_\alpha\| \to 0$, then $|\sup_{\alpha\in A}\|x_{n\alpha}\| - \sup_{\alpha\in A}\|x_\alpha\|| \to 0$). The triangle inequality implies that $\sup_{\alpha\in A}\|x_{n\alpha}\| \leq \sup_{\alpha\in A}\|x_{n\alpha} - x_\alpha\| + \sup_{\alpha\in A}\|x_\alpha\|$, as well as the same result with $x_{n\alpha}$ and x_α interchanged, and the desired result follows from these inequalities. ■

Proof of Lemma 4, further details. It remains to verify the continuity mentioned in the proof. For simplicity, consider the continuity of the functions $g_1 : \mathcal{C}(A,\mathbb{R}^r)\times\{\mathcal{I}_\alpha\} \to \mathcal{B}(A,\mathbb{R})$ mapping $(x_\bullet, \mathcal{I}_\bullet)$ to $x_\bullet'\mathcal{I}_\bullet x_\bullet$ and $g_2 : \mathcal{C}(A,\mathbb{R}^r)\times\{\mathcal{I}_\alpha\} \to \mathcal{B}(A,\mathbb{R})$ mapping $(x_\bullet, \mathcal{I}_\bullet)$ to $\inf_{\boldsymbol{\lambda}\in\Lambda}\{(\boldsymbol{\lambda} - x_\bullet)'\mathcal{I}_\bullet(\boldsymbol{\lambda} - x_\bullet)\}$ separately. For g_1, continuity is rather clear, for if a sequence $(x_{n\bullet}, \mathcal{I}_\bullet)$ converges to $(x_\bullet, \mathcal{I}_\bullet)$ in $\mathcal{C}(A,\mathbb{R}^r)\times\{\mathcal{I}_\alpha\}$, then $g_1((x_{n\bullet}, \mathcal{I}_\bullet)) \to g_1((x_\bullet, \mathcal{I}_\bullet))$ in $\mathcal{B}(A,\mathbb{R})$ (i.e., if $\sup_{\alpha\in A}\|x_{n\alpha} - x_\alpha\| \to 0$, then $\sup_{\alpha\in A}|x_{n\alpha}'\mathcal{I}_\alpha x_{n\alpha} - x_\alpha'\mathcal{I}_\alpha x_\alpha| \to 0$). For the continuity of g_2, suppose that $\sup_{\alpha\in A}\|x_{n\alpha} - x_\alpha\| \to 0$, and consider $\sup_{\alpha\in A}|\inf_{\boldsymbol{\lambda}\in\Lambda}\{(\boldsymbol{\lambda} - x_{n\alpha})'\mathcal{I}_\alpha(\boldsymbol{\lambda} - x_{n\alpha})\} - \inf_{\boldsymbol{\lambda}\in\Lambda}\{(\boldsymbol{\lambda} - x_\alpha)'\mathcal{I}_\alpha(\boldsymbol{\lambda} - x_\alpha)\}|$. Noting that

$$\inf_{\boldsymbol{\lambda}\in\Lambda}\left\{(\boldsymbol{\lambda} - x_{n\alpha})'\mathcal{I}_\alpha(\boldsymbol{\lambda} - x_{n\alpha})\right\} = \left\{\inf_{\boldsymbol{\lambda}\in\Lambda}\|\mathcal{I}_\alpha^{1/2}(\boldsymbol{\lambda} - x_{n\alpha})\|\right\}^2$$

and similarly for the other infimum, we need to consider

$$\sup_{\alpha\in A}\Big\{\Big|\inf_{\boldsymbol{\lambda}\in\Lambda}\|\mathcal{I}_\alpha^{1/2}(\boldsymbol{\lambda}-x_{n\alpha})\|-\inf_{\boldsymbol{\lambda}\in\Lambda}\|\mathcal{I}_\alpha^{1/2}(\boldsymbol{\lambda}-x_\alpha)\|\Big|\Big(\inf_{\boldsymbol{\lambda}\in\Lambda}\|\mathcal{I}_\alpha^{1/2}(\boldsymbol{\lambda}-x_{n\alpha})\|+\inf_{\boldsymbol{\lambda}\in\Lambda}\|\mathcal{I}_\alpha^{1/2}(\boldsymbol{\lambda}-x_\alpha)\|\Big)\Big\}. \tag{43}$$

Using the triangle inequality and properties of the Euclidean vector norm,

$$\|\mathcal{I}_\alpha^{1/2}(\boldsymbol{\lambda}-x_{n\alpha})\|\le\|\mathcal{I}_\alpha^{1/2}(\boldsymbol{\lambda}-x_\alpha)\|+\|\mathcal{I}_\alpha^{1/2}(x_{n\alpha}-x_\alpha)\|\le\|\mathcal{I}_\alpha^{1/2}(\boldsymbol{\lambda}-x_\alpha)\|+\left(\lambda_{\max}\left(\mathcal{I}_\alpha\right)\right)^{1/2}\left\|x_{n\alpha}-x_\alpha\right\|,$$

and similarly with $x_{n\alpha}$ and x_α exchanged, so that

$$\left|\inf_{\boldsymbol{\lambda}\in\Lambda}\|\mathcal{I}_\alpha^{1/2}(\boldsymbol{\lambda}-x_{n\alpha})\|-\inf_{\boldsymbol{\lambda}\in\Lambda}\|\mathcal{I}_\alpha^{1/2}(\boldsymbol{\lambda}-x_\alpha)\|\right|\le\left(\lambda_{\max}\left(\mathcal{I}_\alpha\right)\right)^{1/2}\left\|x_{n\alpha}-x_\alpha\right\|.$$

As was noted after Assumption 6, the cone Λ contains the origin, so that the term in (43) in parentheses is dominated by $\left(\lambda_{\max}\left(\mathcal{I}_\alpha\right)\right)^{1/2}\left(\|x_{n\alpha}\|+\|x_\alpha\|\right)$. Now, due to Assumption 5(iii), the fact that $x_{n\bullet}, x_\bullet$ are bounded, and the assumed $\sup_{\alpha\in A}\|x_{n\alpha}-x_\alpha\|\to 0$, the quantity in (43) converges to zero. ■

Proof of Lemma 5. For brevity and clarity, within this proof we use somewhat simplified notation and let

$$\mathcal{I}_\alpha^{-1}=\begin{bmatrix}A & B\\ B' & C\end{bmatrix}$$

denote the partition of $\mathcal{I}_\alpha^{-1}$ (so that, e.g., C is shorthand for $(\mathcal{I}_\alpha^{-1})_{\vartheta\vartheta}$). This implies that $\mathcal{I}_\alpha$ can be expressed as

$$\mathcal{I}_\alpha=\begin{bmatrix}D^{-1} & -D^{-1}BC^{-1}\\ -C^{-1}B'D^{-1} & C^{-1}+C^{-1}B'D^{-1}BC^{-1}\end{bmatrix}$$

where $D=A-BC^{-1}B'$ (thus, e.g., $D^{-1}=\mathcal{I}_{\theta\theta\alpha}$). Note also that A, C, and D are symmetric (as $\mathcal{I}_\alpha$ is symmetric).

First note that $S_\alpha=\mathcal{I}_\alpha Z_\alpha$ can be expressed as

$$S_\alpha=\begin{bmatrix}D^{-1} & -D^{-1}BC^{-1}\\ -C^{-1}B'D^{-1} & C^{-1}+C^{-1}B'D^{-1}BC^{-1}\end{bmatrix}\begin{bmatrix}Z_{\theta\alpha}\\ Z_{\vartheta\alpha}\end{bmatrix}=\begin{bmatrix}D^{-1}Z_{\theta\alpha}-D^{-1}BC^{-1}Z_{\vartheta\alpha}\\ -C^{-1}B'D^{-1}Z_{\theta\alpha}+C^{-1}Z_{\vartheta\alpha}+C^{-1}B'D^{-1}BC^{-1}Z_{\vartheta\alpha}\end{bmatrix}$$

so that $S_{\theta\alpha}'DS_{\theta\alpha}$ equals

$$\begin{aligned}S_{\theta\alpha}'DS_{\theta\alpha}&=(D^{-1}Z_{\theta\alpha}-D^{-1}BC^{-1}Z_{\vartheta\alpha})'D(D^{-1}Z_{\theta\alpha}-D^{-1}BC^{-1}Z_{\vartheta\alpha})\\&=Z_{\theta\alpha}'D^{-1}Z_{\theta\alpha}-Z_{\theta\alpha}'D^{-1}BC^{-1}Z_{\vartheta\alpha}-Z_{\vartheta\alpha}'C^{-1}B'D^{-1}Z_{\theta\alpha}+Z_{\vartheta\alpha}'C^{-1}B'D^{-1}BC^{-1}Z_{\vartheta\alpha}.\end{aligned}$$

Now, since $Z_\alpha'\mathcal{I}_\alpha Z_\alpha$ can be written as

$$\begin{aligned}Z_\alpha'\mathcal{I}_\alpha Z_\alpha&=\begin{bmatrix}Z_{\theta\alpha}\\ Z_{\vartheta\alpha}\end{bmatrix}'\begin{bmatrix}D^{-1} & -D^{-1}BC^{-1}\\ -C^{-1}B'D^{-1} & C^{-1}+C^{-1}B'D^{-1}BC^{-1}\end{bmatrix}\begin{bmatrix}Z_{\theta\alpha}\\ Z_{\vartheta\alpha}\end{bmatrix}\\&=Z_{\theta\alpha}'D^{-1}Z_{\theta\alpha}-Z_{\theta\alpha}'D^{-1}BC^{-1}Z_{\vartheta\alpha}-Z_{\vartheta\alpha}'C^{-1}B'D^{-1}Z_{\theta\alpha}\\&\quad+Z_{\vartheta\alpha}'C^{-1}Z_{\vartheta\alpha}+Z_{\vartheta\alpha}'C^{-1}B'D^{-1}BC^{-1}Z_{\vartheta\alpha},\end{aligned}$$

we obtain

$$Z_\alpha'\mathcal{I}_\alpha Z_\alpha=Z_{\vartheta\alpha}'C^{-1}Z_{\vartheta\alpha}+S_{\theta\alpha}'DS_{\theta\alpha}. \tag{44}$$

Now consider $\inf_{\boldsymbol{\lambda}\in\Lambda}\{(\boldsymbol{\lambda}-Z_\alpha)'\mathcal{I}_\alpha(\boldsymbol{\lambda}-Z_\alpha)\}$. Similarly as above,

$$
\begin{aligned}
&(\boldsymbol{\lambda}-Z_\alpha)'\mathcal{I}_\alpha(\boldsymbol{\lambda}-Z_\alpha)\\
&=(\boldsymbol{\lambda}_\theta-Z_{\theta\alpha})'D^{-1}(\boldsymbol{\lambda}_\theta-Z_{\theta\alpha})-(\boldsymbol{\lambda}_\theta-Z_{\theta\alpha})'D^{-1}BC^{-1}(\boldsymbol{\lambda}_\vartheta-Z_{\vartheta\alpha})-(\boldsymbol{\lambda}_\vartheta-Z_{\vartheta\alpha})'C^{-1}B'D^{-1}(\boldsymbol{\lambda}_\theta-Z_{\theta\alpha})\\
&+(\boldsymbol{\lambda}_\vartheta-Z_{\vartheta\alpha})'C^{-1}(\boldsymbol{\lambda}_\vartheta-Z_{\vartheta\alpha})+(\boldsymbol{\lambda}_\vartheta-Z_{\vartheta\alpha})'C^{-1}B'D^{-1}BC^{-1}(\boldsymbol{\lambda}_\vartheta-Z_{\vartheta\alpha})\\
&=(\boldsymbol{\lambda}_\theta-Z_{\theta\alpha})'D^{-1}(\boldsymbol{\lambda}_\theta-Z_{\theta\alpha})-(\boldsymbol{\lambda}_\theta-Z_{\theta\alpha})'D^{-1}[BC^{-1}(\boldsymbol{\lambda}_\vartheta-Z_{\vartheta\alpha})]-[BC^{-1}(\boldsymbol{\lambda}_\vartheta-Z_{\vartheta\alpha})]'D^{-1}(\boldsymbol{\lambda}_\theta-Z_{\theta\alpha})\\
&+(\boldsymbol{\lambda}_\vartheta-Z_{\vartheta\alpha})'C^{-1}(\boldsymbol{\lambda}_\vartheta-Z_{\vartheta\alpha})+[BC^{-1}(\boldsymbol{\lambda}_\vartheta-Z_{\vartheta\alpha})]'D^{-1}[BC^{-1}(\boldsymbol{\lambda}_\vartheta-Z_{\vartheta\alpha})]\\
&=(\boldsymbol{\lambda}_\vartheta-Z_{\vartheta\alpha})'C^{-1}(\boldsymbol{\lambda}_\vartheta-Z_{\vartheta\alpha})+\{(\boldsymbol{\lambda}_\theta-Z_{\theta\alpha})-[BC^{-1}(\boldsymbol{\lambda}_\vartheta-Z_{\vartheta\alpha})]\}'D^{-1}\{(\boldsymbol{\lambda}_\theta-Z_{\theta\alpha})-[BC^{-1}(\boldsymbol{\lambda}_\vartheta-Z_{\vartheta\alpha})]\}.
\end{aligned}
$$

Now, for any fixed $\boldsymbol{\lambda}_\vartheta\in\mathbb{R}^{q_\vartheta}$, Assumption 7 implies that

$$\inf_{\boldsymbol{\lambda}_\theta\in\mathbb{R}^{q_\theta}}\{(\boldsymbol{\lambda}_\theta-Z_{\theta\alpha})-[BC^{-1}(\boldsymbol{\lambda}_\vartheta-Z_{\vartheta\alpha})]\}'D^{-1}\{(\boldsymbol{\lambda}_\theta-Z_{\theta\alpha})-[BC^{-1}(\boldsymbol{\lambda}_\vartheta-Z_{\vartheta\alpha})]\}=0$$

(cf. Andrews (1999, eqn. (7.35))) so that

$$\inf_{\boldsymbol{\lambda}\in\Lambda}\left\{(\boldsymbol{\lambda}-Z_\alpha)'\mathcal{I}_\alpha(\boldsymbol{\lambda}-Z_\alpha)\right\}=\inf_{\boldsymbol{\lambda}_\vartheta\in\Lambda_\vartheta}\left\{(\boldsymbol{\lambda}_\vartheta-Z_{\vartheta\alpha})'C^{-1}(\boldsymbol{\lambda}_\vartheta-Z_{\vartheta\alpha})\right\}. \tag{45}$$

Combining (44) and (45) and recalling that $C^{-1}=(\mathcal{I}_\alpha^{-1})_{\vartheta\vartheta}^{-1}$ and $D=\mathcal{I}_{\theta\theta\alpha}^{-1}$ yields the equality stated in the lemma.

Finally, $(\mathcal{I}_\alpha^{-1})_{\vartheta\vartheta}$ and $Z_{\vartheta\alpha}$ can be expressed as

$$
\begin{aligned}
(\mathcal{I}_\alpha^{-1})_{\vartheta\vartheta}&=(\mathcal{I}_{\vartheta\vartheta\alpha}-\mathcal{I}_{\vartheta\theta\alpha}\mathcal{I}_{\theta\theta\alpha}^{-1}\mathcal{I}_{\theta\vartheta\alpha})^{-1}\ [=\mathcal{I}_{\vartheta\vartheta\alpha}^{-1}+\mathcal{I}_{\vartheta\vartheta\alpha}^{-1}\mathcal{I}_{\vartheta\theta\alpha}(\mathcal{I}_{\theta\theta\alpha}-\mathcal{I}_{\theta\vartheta\alpha}\mathcal{I}_{\vartheta\vartheta\alpha}^{-1}\mathcal{I}_{\vartheta\theta\alpha})^{-1}\mathcal{I}_{\theta\vartheta\alpha}\mathcal{I}_{\vartheta\vartheta\alpha}^{-1}],\\
Z_{\vartheta\alpha}&=(\mathcal{I}_\alpha^{-1})_{\vartheta\vartheta}(S_{\vartheta\alpha}-\mathcal{I}_{\vartheta\theta\alpha}\mathcal{I}_{\theta\theta\alpha}^{-1}S_{\theta\alpha})\ [=\mathcal{I}_{\vartheta\vartheta\alpha}^{-1}S_{\vartheta\alpha}+\mathcal{I}_{\vartheta\vartheta\alpha}^{-1}\mathcal{I}_{\vartheta\theta\alpha}(\mathcal{I}_{\theta\theta\alpha}-\mathcal{I}_{\theta\vartheta\alpha}\mathcal{I}_{\vartheta\vartheta\alpha}^{-1}\mathcal{I}_{\vartheta\theta\alpha})^{-1}(\mathcal{I}_{\theta\vartheta\alpha}\mathcal{I}_{\vartheta\vartheta\alpha}^{-1}S_{\vartheta\alpha}-S_{\theta\alpha})],
\end{aligned}
$$

where the two different expressions result from two different ways of writing the inverse of a partitioned matrix. ■

E Further details for the LMAR example

E.1 Verification of Assumption 5(ii), further details

As for the weak convergence requirement in part (ii), we rely on Theorem 2 (and the remarks that follow it) in Andrews and Ploberger (1995). As can be seen from the proof of their Theorem 2, it suffices to verify their conditions EP1(a), EP1(e), and EP4 (omitting the weakly exogeneous X_t variables therein). Under the null hypothesis, y_t is a linear Gaussian AR(p) process so that condition EP1(a) is satisfied with geometrically declining mixing numbers. To check condition EP1(e), we show that $E\big[\sup_{\alpha\in A}\sup_{(\pi,\varpi)\in\Pi}|l_t^{\pi}(\alpha,\pi,\varpi)|\big]<\infty$, $E\big[\sup_{\alpha\in A}\|\nabla_{(\pi,\varpi)}l_t^{\pi}(\alpha,\pi^*,0)\|^r\big]<\infty$ for any positive r, and $E\big[\sup_{\alpha\in A}\sup_{(\pi,\varpi)\in\Pi}\|\nabla^2_{(\pi,\varpi)(\pi,\varpi)'}l_t^{\pi}(\alpha,\pi,\varpi)\|\big]<\infty$. The first of these moment conditions follows from the arguments used to verify our Assumption 2 (see the verification of this Assumption for the GMAR model; the details for the LMAR model are presented there). The second holds due to the expression of $\nabla_{(\pi,\varpi)}l_t^{\pi}(\alpha,\pi^*,0)$ in (27), the fact that $0<\alpha_{1,t}^L(\alpha)<1$, and Lemma F.1. The third is verified below in Supplementary Appedix E.2. As the compactness requirement of condition EP4(a) holds by our Assumption 1(ii), it remains to verify EP4(b). To this end, note that for arbitrary $a,b\in A$,

$$\|\nabla_{(\pi,\varpi)}l_t^{\pi}(a,\pi^*,0)-\nabla_{(\pi,\varpi)}l_t^{\pi}(b,\pi^*,0)\|^r=|\alpha_{1,t}^L(a)-\alpha_{1,t}^L(b)|^r\left\|\frac{\nabla f_t(\pi^*)}{f_t(\pi^*)}\right\|^r.$$

By straightforward differentiation, $\nabla_\alpha\alpha_{1,t}^L(\alpha)=\alpha_{1,t}^L(\alpha)(1-\alpha_{1,t}^L(\alpha))(1,y_{t-1},\ldots,y_{t-m})$, so that by the mean value theorem

$$\alpha_{1,t}^L(a)-\alpha_{1,t}^L(b)=\alpha_{1,t}^L(c_{a,b})(1-\alpha_{1,t}^L(c_{a,b}))(1,y_{t-1},\ldots,y_{t-m})'(a-b)$$

for some $c_{a,b}\in\mathbb{R}^{m+1}$ between a and b (as A is not necessarily convex, $c_{a,b}$ does not necessarily belong to A, but this has no effect in what follows as the expression $\alpha_{1,t}^L(c_{a,b})$ is nevertheless well defined for all $c_{a,b}\in\mathbb{R}^{m+1}$). Setting $B_t=1+|y_{t-1}|+\ldots+|y_{t-m}|$ and noting that $0<\alpha_{1,t}^L(c_{a,b})<1$ this implies that

$$|\alpha_{1,t}^L(a)-\alpha_{1,t}^L(b)|\leq|(1,y_{t-1},\ldots,y_{t-m})'(a-b)|\leq(1+|y_{t-1}|+\ldots+|y_{t-m}|)\|a-b\|=B_t\|a-b\|.$$

Hence

$$E\Big[\sup_{a,b\in A,\|a-b\|<\delta}\|\nabla_{(\pi,\varpi)}l_t^{\pi}(a,\pi^*,0)-\nabla_{(\pi,\varpi)}l_t^{\pi}(b,\pi^*,0)\|^r\Big]<\delta^rE\left[B_t^r\left\|\frac{\nabla f_t(\pi^*)}{f_t(\pi^*)}\right\|^r\right]$$

where on the majorant side the expectation is finite (due to the fact that the y_t's possess moments of all orders, see also the proof of Lemma F.1). Hence condition EP4(b) holds, and the desired weak convergence follows.

E.2 Verification of Assumption 5(iv), further details

It remains to show that $E\big[\sup_{\alpha\in A}\sup_{(\pi,\varpi)\in\Pi}\|\nabla^2_{(\pi,\varpi)(\pi,\varpi)'}l_t^{\pi}(\alpha,\pi,\varpi)\|\big]<\infty$. This in turn follows if we show the same with $\nabla^2_{(\pi,\varpi)(\pi,\varpi)'}l_t^{\pi}(\alpha,\pi,\varpi)$ replaced by $\nabla^2_{\pi\pi'}l_t^{\pi}(\alpha,\pi,\varpi)$, $\nabla^2_{\varpi\varpi'}l_t^{\pi}(\alpha,\pi,\varpi)$, and $\nabla^2_{\pi\varpi'}l_t^{\pi}(\alpha,\pi,\varpi)$. Consider the expression of $\nabla^2_{\pi\pi'}l_t^{\pi}(\alpha,\pi,\varpi)$ given in Appendix B and recall that $0<\alpha_{1,t}^L(\alpha),\alpha_{2,t}^L(\alpha)<1$ and $\nabla f_t(\pi)=f_t(\pi)\nabla_\pi l_t^0(\pi)$ with $l_t^0(\pi)=\log[f_t(\pi)]$. Then we can, for instance, write

$$\left\| \frac{\alpha_{1,t}^{L}(\alpha)\nabla f_t(\pi)}{f_{2,t}^{\pi}(\alpha,\pi,\varpi)} \right\| = \left\| \frac{\alpha_{1,t}^{L}(\alpha) f_t(\pi)\nabla_\pi l_t^0(\pi)}{\alpha_{1,t}^{L}(\alpha)f_t(\pi)+\alpha_{2,t}^{L}(\alpha)f_t(\pi-\varpi)} \right\| \le \left\|\nabla_\pi l_t^0(\pi)\right\|$$

and

$$\begin{aligned}\left\| \frac{\alpha_{1,t}^{L}(\alpha)\nabla' f_t(\pi)+\alpha_{2,t}^{L}(\alpha)\nabla' f_t(\pi-\varpi)}{f_{2,t}^{\pi}(\alpha,\pi,\varpi)} \right\| &= \left\| \frac{\alpha_{1,t}^{L}(\alpha) f_t(\pi)\nabla_\pi l_t^0(\pi)+\alpha_{2,t}^{L}(\alpha) f_t(\pi-\varpi)\nabla_\pi l_t^0(\pi-\varpi)}{\alpha_{1,t}^{L}(\alpha)f_t(\pi)+\alpha_{2,t}^{L}(\alpha)f_t(\pi-\varpi)} \right\| \\ &\le \left\|\nabla_\pi l_t^0(\pi)\right\| + \left\|\nabla_\pi l_t^0(\pi-\varpi)\right\|.\end{aligned}$$

As similar inequalities can be obtained for the second term of $\nabla^2_{\pi\pi'} l_t^\pi(\alpha,\pi,\varpi)$, we get

$$\begin{aligned}\left\|\nabla^2_{\pi\pi'} l_t^\pi(\alpha,\pi,\varpi)\right\| \le{}& \left\|\alpha_{1,t}^{L}(\alpha)\frac{\nabla^2 f_t(\pi)}{f_{2,t}^{\pi}(\alpha,\pi,\varpi)}\right\| + \left\|\nabla_\pi l_t^0(\pi)\right\|^2 + \left\|\nabla_\pi l_t^0(\pi)\right\| \left\|\nabla_\pi l_t^0(\pi-\varpi)\right\| \\ &+ \left\|\alpha_{2,t}^{L}(\alpha)\frac{\nabla^2 f_t(\pi-\varpi)}{f_{2,t}^{\pi}(\alpha,\pi,\varpi)}\right\| + \left\|\nabla_\pi l_t^0(\pi-\varpi)\right\|^2 + \left\|\nabla_\pi l_t^0(\pi)\right\| \left\|\nabla_\pi l_t^0(\pi-\varpi)\right\|.\end{aligned}$$

Next note that

$$\frac{\nabla^2 f_t(\pi)}{f_{2,t}^{\pi}(\alpha,\pi,\varpi)} = \frac{\nabla (f_t(\pi)\nabla_\pi l_t^0(\pi))}{f_{2,t}^{\pi}(\alpha,\pi,\varpi)} = \frac{f_t(\pi)\nabla_\pi l_t^0(\pi)\nabla_{\pi'} l_t^0(\pi)}{f_{2,t}^{\pi}(\alpha,\pi,\varpi)} + \frac{f_t(\pi)\nabla^2_{\pi\pi'} l_t^0(\pi)}{f_{2,t}^{\pi}(\alpha,\pi,\varpi)},$$

so that arguments similar to those already used above give

$$\left\|\alpha_{1,t}^{L}(\alpha)\frac{\nabla^2 f_t(\pi)}{f_{2,t}^{\pi}(\alpha,\pi,\varpi)}\right\| \le \left\|\nabla_\pi l_t^0(\pi)\right\|^2 + \left\|\nabla^2_{\pi\pi'} l_t^0(\pi)\right\|$$

and

$$\left\|\alpha_{2,t}^{L}(\alpha)\frac{\nabla^2 f_t(\pi-\varpi)}{f_{2,t}^{\pi}(\alpha,\pi,\varpi)}\right\| \le \left\|\nabla_\pi l_t^0(\pi-\varpi)\right\|^2 + \left\|\nabla^2_{\pi\pi'} l_t^0(\pi-\varpi)\right\|.$$

Hence, we can conclude that

$$\begin{aligned}\left\|\nabla^2_{\pi\pi'} l_t^\pi(\alpha,\pi,\varpi)\right\| \le{}& 2\left\|\nabla_\pi l_t^0(\pi)\right\|^2 + 2\left\|\nabla_\pi l_t^0(\pi-\varpi)\right\|^2 + 2\left\|\nabla_\pi l_t^0(\pi)\right\| \left\|\nabla_\pi l_t^0(\pi-\varpi)\right\| \\ &+ \left\|\nabla^2_{\pi\pi'} l_t^0(\pi)\right\| + \left\|\nabla^2_{\pi\pi'} l_t^0(\pi-\varpi)\right\|.\end{aligned}$$

To bound the expression on the dominant side, note that $\nabla_\pi l_t^0(\pi) = \frac{\nabla f_t(\pi)}{f_t(\pi)}$ and $\nabla^2_{\pi\pi'} l_t^0(\pi) = \frac{\nabla^2 f_t(\pi)}{f_t(\pi)} - \frac{\nabla f_t(\pi)}{f_t(\pi)}\frac{\nabla' f_t(\pi)}{f_t(\pi)}$ so that Lemma F.1 ensures that $E\left[\sup_{\alpha\in A}\sup_{(\pi,\varpi)\in\Pi}\left\|\nabla^2_{\pi\pi'} l_t^\pi(\alpha,\pi,\varpi)\right\|\right] < \infty$. An inspection of the expressions of $\nabla^2_{\pi\varpi'} l_t^\pi(\alpha,\pi,\varpi)$ and $\nabla^2_{\varpi\varpi'} l_t^\pi(\alpha,\pi,\varpi)$ in Appendix B shows that a similar result can be obtained with $\nabla^2_{\pi\pi'} l_t^\pi(\alpha,\pi,\varpi)$ replaced by $\nabla^2_{\pi\varpi'} l_t^\pi(\alpha,\pi,\varpi)$ and $\nabla^2_{\varpi\varpi'} l_t^\pi(\alpha,\pi,\varpi)$, yielding the desired result.

F Further details for the GMAR example

F.1 Partial derivatives of the reparameterized log-likelihood function

Here we present certain partial derivatives of $l_t^{\pi}(\alpha,\beta,\pi,\varpi)$ with respect to (β,π,ϖ). For brevity, set $\tilde{\pi}=(\beta,\pi)$ (and similarly $\tilde{\pi}^*=(\beta^*,\pi^*)$), so that the desired derivatives are with respect to $\tilde{\pi}$ and ϖ or, elementwise, with respect to $\tilde{\pi}_i$ and ϖ_j for $i=1,\ldots,p+2$ and $j=1,\ldots,p+1$. In the derivative expressions below, the subindices in $\tilde{\pi}$ and ϖ are tacitly assumed to be withing these ranges. For brevity, denote $l_t^{\pi*}=l_t^{\pi}(\alpha,\tilde{\pi}^*,0)$, $f_t^*=f_t(\tilde{\pi}^*)$, $\mathsf{n}_p^*=\mathsf{n}_p(\tilde{\pi}^*)$, and similarly for their partial derivatives.

The following derivatives are obtained with straightforward (but tedious and lengthy) differentiation. The necessary calculations for the first- and second-order derivatives are presented in Supplementary Appendix F.7, but for brevity we omit the detailed calculations for the third- and fourth-order derivatives.

First- and second-order derivatives:

$$
\begin{aligned}
\nabla_{\tilde{\pi}_i} l_t^{\pi*} &= \frac{\nabla_i f_t^*}{f_t^*} \\
\nabla_{\varpi_j} l_t^{\pi*} &= 0 \\
\nabla^2_{\tilde{\pi}_i\tilde{\pi}_j} l_t^{\pi*} &= \frac{\nabla^2_{ij} f_t^*}{f_t^*} - \frac{\nabla_i f_t^*}{f_t^*}\frac{\nabla_j f_t^*}{f_t^*}. \\
\nabla^2_{\tilde{\pi}_i\varpi_j} l_t^{\pi*} &= 0 \\
\nabla^2_{\varpi_i\varpi_j} l_t^{\pi*} &= \alpha_1\alpha_2\left[\frac{\nabla_{i+1}\mathsf{n}_p^*}{\mathsf{n}_p^*}\frac{\nabla_{j+1} f_t^*}{f_t^*} + \frac{\nabla_{i+1} f_t^*}{f_t^*}\frac{\nabla_{j+1}\mathsf{n}_p^*}{\mathsf{n}_p^*} + \frac{\nabla^2_{i+1,j+1} f_t^*}{f_t^*}\right]
\end{aligned}
$$

Third-order derivatives:

$$
\begin{aligned}
\nabla^3_{\tilde{\pi}_i\tilde{\pi}_j\tilde{\pi}_k} l_t^{\pi*} &= \frac{\nabla^3_{ijk} f_t^*}{f_t^*} - \frac{\nabla^2_{ij} f_t^*}{f_t^*}\frac{\nabla_k f_t^*}{f_t^*} - \frac{\nabla^2_{ik} f_t^*}{f_t^*}\frac{\nabla_j f_t^*}{f_t^*} - \frac{\nabla^2_{jk} f_t^*}{f_t^*}\frac{\nabla_i f_t^*}{f_t^*} + 2\frac{\nabla_i f_t^*}{f_t^*}\frac{\nabla_j f_t^*}{f_t^*}\frac{\nabla_k f_t^*}{f_t^*} \\
\nabla^3_{\tilde{\pi}_i\tilde{\pi}_j\varpi_k} l_t^{\pi*} &= -\alpha_1\alpha_2\frac{\nabla_i f_t^*}{f_t^*}\left(\frac{\nabla^2_{j,k+1}\mathsf{n}_p^*}{\mathsf{n}_p^*} - \frac{\nabla_j\mathsf{n}_p^*}{\mathsf{n}_p^*}\frac{\nabla_{k+1}\mathsf{n}_p^*}{\mathsf{n}_p^*}\right) \\
\nabla^3_{\tilde{\pi}_i\varpi_j\varpi_k} l_t^{\pi*} &= \alpha_1\alpha_2\Bigg[\frac{\nabla_{k+1}\mathsf{n}_p^*}{\mathsf{n}_p^*}\left(\frac{\nabla^2_{i,j+1} f_t^*}{f_t^*} - \frac{\nabla_i f_t^*}{f_t^*}\frac{\nabla_{j+1} f_t^*}{f_t^*}\right) + \frac{\nabla_{j+1}\mathsf{n}_p^*}{\mathsf{n}_p^*}\left(\frac{\nabla^2_{i,k+1} f_t^*}{f_t^*} - \frac{\nabla_i f_t^*}{f_t^*}\frac{\nabla_{k+1} f_t^*}{f_t^*}\right) \\
&\quad + \left(\frac{\nabla^3_{i,j+1,k+1} f_t^*}{f_t^*} - \frac{\nabla_i f_t^*}{f_t^*}\frac{\nabla^2_{j+1,k+1} f_t^*}{f_t^*}\right) + \left(\frac{\nabla^2_{i,k+1}\mathsf{n}_p^*}{\mathsf{n}_p^*} - \frac{\nabla_i\mathsf{n}_p^*}{\mathsf{n}_p^*}\frac{\nabla_{k+1}\mathsf{n}_p^*}{\mathsf{n}_p^*}\right)\frac{\nabla_{j+1} f_t^*}{f_t^*} \\
&\quad + \left(\frac{\nabla^2_{i,j+1}\mathsf{n}_p^*}{\mathsf{n}_p^*} - \frac{\nabla_i\mathsf{n}_p^*}{\mathsf{n}_p^*}\frac{\nabla_{j+1}\mathsf{n}_p^*}{\mathsf{n}_p^*}\right)\frac{\nabla_{k+1} f_t^*}{f_t^*}\Bigg] \\
\nabla^3_{\varpi_i\varpi_j\varpi_k} l_t^{\pi*} &= \alpha_1\alpha_2(\alpha_2-\alpha_1)\Bigg[\frac{\nabla^2_{i+1,j+1}\mathsf{n}_p^*}{\mathsf{n}_p^*}\frac{\nabla_{k+1} f_t^*}{f_t^*} + \frac{\nabla^2_{i+1,k+1}\mathsf{n}_p^*}{\mathsf{n}_p^*}\frac{\nabla_{j+1} f_t^*}{f_t^*} + \frac{\nabla^2_{j+1,k+1}\mathsf{n}_p^*}{\mathsf{n}_p^*}\frac{\nabla_{i+1} f_t^*}{f_t^*} \\
&\quad + \frac{\nabla_{i+1}\mathsf{n}_p^*}{\mathsf{n}_p^*}\frac{\nabla^2_{j+1,k+1} f_t^*}{f_t^*} + \frac{\nabla_{j+1}\mathsf{n}_p^*}{\mathsf{n}_p^*}\frac{\nabla^2_{i+1,k+1} f_t^*}{f_t^*} + \frac{\nabla_{k+1}\mathsf{n}_p^*}{\mathsf{n}_p^*}\frac{\nabla^2_{i+1,j+1} f_t^*}{f_t^*} \\
&\quad + \frac{\nabla^3_{i+1,j+1,k+1} f_t^*}{f_t^*}\Bigg]
\end{aligned}
$$

Fourth-order derivative (fourth-order derivatives with respect to $\tilde{\pi}$ will not be explicitly needed, and thus we omit their expressions):

$$
\begin{aligned}
&\nabla^4_{\varpi_i\varpi_j\varpi_k\varpi_l} l_t^{\pi*} \\
&= -\alpha_1^2\alpha_2^2 \Bigg[\left(\frac{\nabla_{i+1}\mathsf{n}_p^*}{\mathsf{n}_p^*} \frac{\nabla^2_{j+1,k+1}\mathsf{n}_p^*}{\mathsf{n}_p^*} + \frac{\nabla_{j+1}\mathsf{n}_p^*}{\mathsf{n}_p^*} \frac{\nabla^2_{i+1,k+1}\mathsf{n}_p^*}{\mathsf{n}_p^*} + \frac{\nabla_{k+1}\mathsf{n}_p^*}{\mathsf{n}_p^*} \frac{\nabla^2_{i+1,j+1}\mathsf{n}_p^*}{\mathsf{n}_p^*} \right) \frac{\nabla_{l+1} f_t^*}{f_t^*} \\
&\qquad + \left(\frac{\nabla_{i+1}\mathsf{n}_p^*}{\mathsf{n}_p^*} \frac{\nabla^2_{j+1,l+1}\mathsf{n}_p^*}{\mathsf{n}_p^*} + \frac{\nabla_{j+1}\mathsf{n}_p^*}{\mathsf{n}_p^*} \frac{\nabla^2_{i+1,l+1}\mathsf{n}_p^*}{\mathsf{n}_p^*} + \frac{\nabla_{l+1}\mathsf{n}_p^*}{\mathsf{n}_p^*} \frac{\nabla^2_{i+1,j+1}\mathsf{n}_p^*}{\mathsf{n}_p^*} \right) \frac{\nabla_{k+1} f_t^*}{f_t^*} \\
&\qquad + \left(\frac{\nabla_{i+1}\mathsf{n}_p^*}{\mathsf{n}_p^*} \frac{\nabla^2_{k+1,l+1}\mathsf{n}_p^*}{\mathsf{n}_p^*} + \frac{\nabla_{k+1}\mathsf{n}_p^*}{\mathsf{n}_p^*} \frac{\nabla^2_{i+1,l+1}\mathsf{n}_p^*}{\mathsf{n}_p^*} + \frac{\nabla_{l+1}\mathsf{n}_p^*}{\mathsf{n}_p^*} \frac{\nabla^2_{i+1,k+1}\mathsf{n}_p^*}{\mathsf{n}_p^*} \right) \frac{\nabla_{j+1} f_t^*}{f_t^*} \\
&\qquad + \left(\frac{\nabla_{j+1}\mathsf{n}_p^*}{\mathsf{n}_p^*} \frac{\nabla^2_{k+1,l+1}\mathsf{n}_p^*}{\mathsf{n}_p^*} + \frac{\nabla_{k+1}\mathsf{n}_p^*}{\mathsf{n}_p^*} \frac{\nabla^2_{j+1,l+1}\mathsf{n}_p^*}{\mathsf{n}_p^*} + \frac{\nabla_{l+1}\mathsf{n}_p^*}{\mathsf{n}_p^*} \frac{\nabla^2_{j+1,k+1}\mathsf{n}_p^*}{\mathsf{n}_p^*} \right) \frac{\nabla_{i+1} f_t^*}{f_t^*} \\
&\qquad + \left(\frac{\nabla_{i+1}\mathsf{n}_p^*}{\mathsf{n}_p^*} \frac{\nabla_{j+1} f_t^*}{f_t^*} + \frac{\nabla_{j+1}\mathsf{n}_p^*}{\mathsf{n}_p^*} \frac{\nabla_{i+1} f_t^*}{f_t^*} + \frac{\nabla^2_{i+1,j+1} f_t^*}{f_t^*} \right) \\
&\qquad\qquad \times \left(\frac{\nabla_{k+1}\mathsf{n}_p^*}{\mathsf{n}_p^*} \frac{\nabla_{l+1} f_t^*}{f_t^*} + \frac{\nabla_{l+1}\mathsf{n}_p^*}{\mathsf{n}_p^*} \frac{\nabla_{k+1} f_t^*}{f_t^*} + \frac{\nabla^2_{k+1,l+1} f_t^*}{f_t^*} \right) \\
&\qquad + \left(\frac{\nabla_{i+1}\mathsf{n}_p^*}{\mathsf{n}_p^*} \frac{\nabla_{k+1} f_t^*}{f_t^*} + \frac{\nabla_{k+1}\mathsf{n}_p^*}{\mathsf{n}_p^*} \frac{\nabla_{i+1} f_t^*}{f_t^*} + \frac{\nabla^2_{i+1,k+1} f_t^*}{f_t^*} \right) \\
&\qquad\qquad \times \left(\frac{\nabla_{j+1}\mathsf{n}_p^*}{\mathsf{n}_p^*} \frac{\nabla_{l+1} f_t^*}{f_t^*} + \frac{\nabla_{l+1}\mathsf{n}_p^*}{\mathsf{n}_p^*} \frac{\nabla_{j+1} f_t^*}{f_t^*} + \frac{\nabla^2_{j+1,l+1} f_t^*}{f_t^*} \right) \\
&\qquad + \left(\frac{\nabla_{i+1}\mathsf{n}_p^*}{\mathsf{n}_p^*} \frac{\nabla_{l+1} f_t^*}{f_t^*} + \frac{\nabla_{l+1}\mathsf{n}_p^*}{\mathsf{n}_p^*} \frac{\nabla_{i+1} f_t^*}{f_t^*} + \frac{\nabla^2_{i+1,l+1} f_t^*}{f_t^*} \right) \\
&\qquad\qquad \times \left(\frac{\nabla_{j+1}\mathsf{n}_p^*}{\mathsf{n}_p^*} \frac{\nabla_{k+1} f_t^*}{f_t^*} + \frac{\nabla_{k+1}\mathsf{n}_p^*}{\mathsf{n}_p^*} \frac{\nabla_{j+1} f_t^*}{f_t^*} + \frac{\nabla^2_{j+1,k+1} f_t^*}{f_t^*} \right) \Bigg] \\
&+ \alpha_1\alpha_2(1-3\alpha_1\alpha_2) \Bigg[\frac{\nabla^3_{i+1,j+1,k+1}\mathsf{n}_p^*}{\mathsf{n}_p^*} \frac{\nabla_{l+1} f_t^*}{f_t^*} + \frac{\nabla^3_{i+1,j+1,l+1}\mathsf{n}_p^*}{\mathsf{n}_p^*} \frac{\nabla_{k+1} f_t^*}{f_t^*} + \frac{\nabla^3_{i+1,k+1,l+1}\mathsf{n}_p^*}{\mathsf{n}_p^*} \frac{\nabla_{j+1} f_t^*}{f_t^*} \\
&\qquad\qquad + \frac{\nabla^3_{j+1,k+1,l+1}\mathsf{n}_p^*}{\mathsf{n}_p^*} \frac{\nabla_{i+1} f_t^*}{f_t^*} + \frac{\nabla_{i+1}\mathsf{n}_p^*}{\mathsf{n}_p^*} \frac{\nabla^3_{j+1,k+1,l+1} f_t^*}{f_t^*} + \frac{\nabla_{j+1}\mathsf{n}_p^*}{\mathsf{n}_p^*} \frac{\nabla^3_{i+1,k+1,l+1} f_t^*}{f_t^*} \\
&\qquad\qquad + \frac{\nabla_{k+1}\mathsf{n}_p^*}{\mathsf{n}_p^*} \frac{\nabla^3_{i+1,j+1,l+1} f_t^*}{f_t^*} + \frac{\nabla_{l+1}\mathsf{n}_p^*}{\mathsf{n}_p^*} \frac{\nabla^3_{i+1,j+1,k+1} f_t^*}{f_t^*} + \frac{\nabla^4_{i+1,j+1,k+1,l+1} f_t^*}{f_t^*} \Bigg] \\
&+ \alpha_1\alpha_2(\alpha_2-\alpha_1)^2 \Bigg[\frac{\nabla^2_{i+1,j+1}\mathsf{n}_p^*}{\mathsf{n}_p^*} \frac{\nabla^2_{k+1,l+1} f_t^*}{f_t^*} + \frac{\nabla^2_{i+1,k+1}\mathsf{n}_p^*}{\mathsf{n}_p^*} \frac{\nabla^2_{j+1,l+1} f_t^*}{f_t^*} + \frac{\nabla^2_{i+1,l+1}\mathsf{n}_p^*}{\mathsf{n}_p^*} \frac{\nabla^2_{j+1,k+1} f_t^*}{f_t^*} \\
&\qquad\qquad + \frac{\nabla^2_{j+1,k+1}\mathsf{n}_p^*}{\mathsf{n}_p^*} \frac{\nabla^2_{i+1,l+1} f_t^*}{f_t^*} + \frac{\nabla^2_{j+1,l+1}\mathsf{n}_p^*}{\mathsf{n}_p^*} \frac{\nabla^2_{i+1,k+1} f_t^*}{f_t^*} + \frac{\nabla^2_{k+1,l+1}\mathsf{n}_p^*}{\mathsf{n}_p^*} \frac{\nabla^2_{i+1,j+1} f_t^*}{f_t^*} \Bigg]
\end{aligned}
$$

F.2 Fourth-order expansion of the log-likelihood function

Justification of (36) and expression of $R_T^{(1)}(\alpha, \tilde{\pi}, \varpi)$. Straightforward calculation yields the fourth-order Taylor expansion (36) with the remainder term (for brevity, we again write $\tilde{\pi} = (\beta, \pi)$ and $\tilde{\pi}^* = (\beta^*, \pi^*)$)

$$
\begin{aligned}
R_T^{(1)}(\alpha, \tilde{\pi}, \varpi) &= \varpi' \nabla_{\varpi} L_T^{\pi}(\alpha, \tilde{\pi}^*, 0) + \frac{2}{2!} (\tilde{\pi} - \tilde{\pi}^*)' \nabla^2_{\tilde{\pi}\varpi'} L_T^{\pi}(\alpha, \tilde{\pi}^*, 0) \varpi \\
&+ \frac{1}{3!} \sum_{i=1}^{q_1+q_2} \sum_{j=1}^{q_1+q_2} \sum_{k=1}^{q_1+q_2} \nabla^3_{\tilde{\pi}_i \tilde{\pi}_j \tilde{\pi}_k} L_T^{\pi}(\alpha, \tilde{\pi}^*, 0) (\tilde{\pi}_i - \tilde{\pi}_i^*)(\tilde{\pi}_j - \tilde{\pi}_j^*)(\tilde{\pi}_k - \tilde{\pi}_k^*) \\
&+ \frac{3}{3!} \sum_{i=1}^{q_1+q_2} \sum_{j=1}^{q_1+q_2} \sum_{k=1}^{q_2} \nabla^3_{\tilde{\pi}_i \tilde{\pi}_j \varpi_k} L_T^{\pi}(\alpha, \tilde{\pi}^*, 0) (\tilde{\pi}_i - \tilde{\pi}_i^*)(\tilde{\pi}_j - \tilde{\pi}_j^*) \varpi_k \\
&+ \frac{1}{3!} \sum_{i=1}^{q_2} \sum_{j=1}^{q_2} \sum_{k=1}^{q_2} \nabla^3_{\varpi_i \varpi_j \varpi_k} L_T^{\pi}(\alpha, \tilde{\pi}^*, 0) \varpi_i \varpi_j \varpi_k \\
&+ \frac{1}{4!} \sum_{i=1}^{q_1+q_2} \sum_{j=1}^{q_1+q_2} \sum_{k=1}^{q_1+q_2} \sum_{l=1}^{q_1+q_2} \nabla^4_{\tilde{\pi}_i \tilde{\pi}_j \tilde{\pi}_k \tilde{\pi}_l} L_T^{\pi}(\alpha, \dot{\tilde{\pi}}, \dot{\varpi}) (\tilde{\pi}_i - \tilde{\pi}_i^*)(\tilde{\pi}_j - \tilde{\pi}_j^*)(\tilde{\pi}_k - \tilde{\pi}_k^*)(\tilde{\pi}_l - \tilde{\pi}_l^*) \\
&+ \frac{4}{4!} \sum_{i=1}^{q_1+q_2} \sum_{j=1}^{q_1+q_2} \sum_{k=1}^{q_1+q_2} \sum_{l=1}^{q_2} \nabla^4_{\tilde{\pi}_i \tilde{\pi}_j \tilde{\pi}_k \varpi_l} L_T^{\pi}(\alpha, \dot{\tilde{\pi}}, \dot{\varpi}) (\tilde{\pi}_i - \tilde{\pi}_i^*)(\tilde{\pi}_j - \tilde{\pi}_j^*)(\tilde{\pi}_k - \tilde{\pi}_k^*) \varpi_l \\
&+ \frac{6}{4!} \sum_{i=1}^{q_1+q_2} \sum_{j=1}^{q_1+q_2} \sum_{k=1}^{q_2} \sum_{l=1}^{q_2} \nabla^4_{\tilde{\pi}_i \tilde{\pi}_j \varpi_k \varpi_l} L_T^{\pi}(\alpha, \dot{\tilde{\pi}}, \dot{\varpi}) (\tilde{\pi}_i - \tilde{\pi}_i^*)(\tilde{\pi}_j - \tilde{\pi}_j^*) \varpi_k \varpi_l \\
&+ \frac{4}{4!} \sum_{i=1}^{q_1+q_2} \sum_{j=1}^{q_2} \sum_{k=1}^{q_2} \sum_{l=1}^{q_2} \nabla^4_{\tilde{\pi}_i \varpi_j \varpi_k \varpi_l} L_T^{\pi}(\alpha, \dot{\tilde{\pi}}, \dot{\varpi}) (\tilde{\pi}_i - \tilde{\pi}_i^*) \varpi_j \varpi_k \varpi_l \\
&+ \frac{1}{4!} \sum_{i=1}^{q_2} \sum_{j=1}^{q_2} \sum_{k=1}^{q_2} \sum_{l=1}^{q_2} \big(\nabla^4_{\varpi_i \varpi_j \varpi_k \varpi_l} L_T^{\pi}(\alpha, \dot{\tilde{\pi}}, \dot{\varpi}) - \nabla_{\varpi_i \varpi_j \varpi_k \varpi_l} L_T^{\pi}(\alpha, \tilde{\pi}^*, 0) \big) \varpi_i \varpi_j \varpi_k \varpi_l, \quad (46)
\end{aligned}
$$

where $(\dot{\tilde{\pi}}, \dot{\varpi})$ denotes a point between $(\tilde{\pi}, \varpi)$ and $(\tilde{\pi}^*, 0)$.

Justification of (37) and expression of $R_T(\alpha, \tilde{\pi}, \varpi)$. We begin with some useful notation. Let $\mathfrak{I}$ denote the index set

$$\mathfrak{I} = ((1,1), (2,2), \ldots, (q_2, q_2), (1,2), (1,3), \ldots, (1, q_2), (2,3), \ldots, (q_2 - 1, q_2)).$$

For any scalars (or $d \times 1$ column vectors) A_{ij} indexed by i and j (here and elsewhere it is tacitly assumed these indices belong to $\{1, \ldots, q_2\}$), let $[A_{ij}]_{(i,j) \in \mathfrak{I}}$ denote the following $1 \times q_2(q_2+1)/2$ row vector (or $d \times q_2(q_2+1)/2$ matrix):

$$[A_{ij}]_{(i,j) \in \mathfrak{I}} = [A_{11} : \cdots : A_{q_2 q_2} : A_{12} : \cdots : A_{q_2 - 1, q_2}] \,.$$

For instance, $v(\varpi) = [\varpi_i \varpi_j]'_{(i,j) \in \mathfrak{I}}$. Similarly, for any scalars A_{ijkl} indexed by i, j, k, l, let $[A_{ijkl}]_{(i,j,k,l) \in \mathfrak{I} \times \mathfrak{I}}$

denote the following $q_2(q_2+1)/2 \times q_2(q_2+1)/2$ matrix

$$[A_{ijkl}]_{(i,j,k,l)\in\mathfrak{I}\times\mathfrak{I}} = \begin{bmatrix} A_{1111} & \cdots & A_{11q_2q_2} & A_{1112} & \cdots & A_{1,1,q_2-1,q_2} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ A_{q_2q_211} & \cdots & A_{q_2q_2q_2q_2} & A_{q_2q_212} & \cdots & A_{q_2,q_2,q_2-1,q_2} \\ A_{1211} & \cdots & A_{12q_2q_2} & A_{1212} & \cdots & A_{1,2,q_2-1,q_2} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ A_{q_2-1,q_2,1,1} & \cdots & A_{q_2-1,q_2,q_2,q_2} & A_{q_2-1,q_2,1,2} & \cdots & A_{q_2-1,q_2,q_2-1,q_2} \end{bmatrix}.$$

With this notation, and for any scalars A_{ijkl} and B_{ij} such that $A_{ijkl} = A_{jikl}$, $A_{ijkl} = A_{ijlk}$, and $B_{ij} = B_{ji}$ for all i,j,k,l, it holds that[12]

$$[B_{ij}]_{(i,j)\in\mathfrak{I}} [c_{ij}c_{kl}A_{ijkl}]_{(i,j,k,l)\in\mathfrak{I}\times\mathfrak{I}} [B_{kl}]'_{(k,l)\in\mathfrak{I}} = \frac{1}{4}\sum_{i=1}^{q_2}\sum_{j=1}^{q_2}\sum_{k=1}^{q_2}\sum_{l=1}^{q_2} A_{ijkl}B_{ij}B_{kl}, \tag{47}$$

where the c_{ij}'s are as in Section 3.3.1 ($c_{ij} = 1/2$ if $i = j$ and $c_{ij} = 1$ if $i \neq j$).

Now, to obtain (37), introduce the matrix

$$\mathcal{J}_T = \begin{bmatrix} \mathcal{J}_{T,\tilde{\pi}\tilde{\pi}} & \mathcal{J}'_{T,\tilde{\pi}\varpi\varpi} \\ \mathcal{J}_{T,\tilde{\pi}\varpi\varpi} & \mathcal{J}_{T,\varpi\varpi\varpi\varpi} \end{bmatrix}$$

where the matrices $\mathcal{J}_{T,\tilde{\pi}\tilde{\pi}}$ $((q_1+q_2)\times(q_1+q_2))$, $\mathcal{J}'_{T,\tilde{\pi}\varpi\varpi}$ $((q_1+q_2)\times q_\vartheta)$, and $\mathcal{J}_{T,\varpi\varpi\varpi\varpi}$ $(q_\vartheta \times q_\vartheta)$ are defined as follows (here $\nabla^2_{\tilde{\pi}\tilde{\pi}'}L_T^{\pi*}$ stands for $\nabla^2_{\tilde{\pi}\tilde{\pi}'}L_T^{\pi}(\alpha,\tilde{\pi}^*,0)$ etc.)

$$\begin{aligned} \mathcal{J}_{T,\tilde{\pi}\tilde{\pi}} &= -T^{-1}\nabla^2_{\tilde{\pi}\tilde{\pi}'}L_T^{\pi*}, \\ \mathcal{J}'_{T,\tilde{\pi}\varpi\varpi} &= -T^{-1}\frac{1}{\alpha_1\alpha_2}\left[c_{ij}\nabla^3_{\tilde{\pi}\varpi_i\varpi_j}L_T^{\pi*}\right]_{(i,j)\in\mathfrak{I}} \\ &= -T^{-1}\frac{1}{\alpha_1\alpha_2}[c_{11}\nabla^3_{\tilde{\pi}\varpi_1\varpi_1}L_T^{\pi*} : \cdots : c_{q_2-1,q_2}\nabla^3_{\tilde{\pi}\varpi_{q_2-1}\varpi_{q_2}}L_T^{\pi*}], \\ \mathcal{J}_{T,\varpi\varpi\varpi\varpi} &= -T^{-1}\frac{8}{4!}\frac{1}{\alpha_1^2\alpha_2^2}\left[c_{ij}c_{kl}\nabla^4_{\varpi_i\varpi_j\varpi_k\varpi_l}L_T^{\pi*}\right]_{(i,j,k,l)\in\mathfrak{I}\times\mathfrak{I}} \\ &= -T^{-1}\frac{8}{4!}\frac{1}{\alpha_1^2\alpha_2^2}\begin{bmatrix} c_{11}c_{11}\nabla^4_{\varpi_1\varpi_1\varpi_1\varpi_1}L_T^{\pi*} & \cdots & c_{q_2-1,q_2}c_{11}\nabla^4_{\varpi_{q_2-1}\varpi_{q_2}\varpi_1\varpi_1}L_T^{\pi*} \\ \vdots & \ddots & \vdots \\ c_{11}c_{q_2-1,q_2}\nabla^4_{\varpi_1\varpi_1\varpi_{q_2-1}\varpi_{q_2}}L_T^{\pi*} & \cdots & c_{q_2-1,q_2}c_{q_2-1,q_2}\nabla^4_{\varpi_{q_2-1}\varpi_{q_2}\varpi_{q_2-1}\varpi_{q_2}}L_T^{\pi*} \end{bmatrix}. \end{aligned}$$

[12] To justify (47), partition the index set $\mathfrak{I}$ as $\mathfrak{I} = (\mathfrak{I}_1, \mathfrak{I}_2)$ with $\mathfrak{I}_1 = ((1,1),(2,2),\ldots(q_2,q_2))$ and $\mathfrak{I}_2 = ((1,2),(1,3),\ldots,(1,q_2),(2,3),\ldots,(q_2-1,q_2))$. With straightforward algebra,

$$\begin{aligned} &[B_{ij}]_{(i,j)\in\mathfrak{I}} [c_{ij}c_{kl}A_{ijkl}]_{(i,j,k,l)\in\mathfrak{I}\times\mathfrak{I}} [B_{kl}]'_{(k,l)\in\mathfrak{I}} \\ &= \sum_{(i,j)\in\mathfrak{I}}\sum_{(k,l)\in\mathfrak{I}} c_{ij}c_{kl}A_{ijkl}B_{ij}B_{kl} \\ &= \frac{1}{4}\sum_{(i,j)\in\mathfrak{I}_1}\sum_{(k,l)\in\mathfrak{I}_1} A_{ijkl}B_{ij}B_{kl} + \frac{1}{2}\sum_{(i,j)\in\mathfrak{I}_1}\sum_{(k,l)\in\mathfrak{I}_2} A_{ijkl}B_{ij}B_{kl} + \frac{1}{2}\sum_{(i,j)\in\mathfrak{I}_2}\sum_{(k,l)\in\mathfrak{I}_1} A_{ijkl}B_{ij}B_{kl} + \sum_{(i,j)\in\mathfrak{I}_2}\sum_{(k,l)\in\mathfrak{I}_2} A_{ijkl}B_{ij}B_{kl} \\ &= \frac{1}{4}\left[\sum_{i=j}\sum_{k=l} A_{ijkl}B_{ij}B_{kl} + 2\sum_{i=j}\sum_{k<l} A_{ijkl}B_{ij}B_{kl} + 2\sum_{i<j}\sum_{k=l} A_{ijkl}B_{ij}B_{kl} + 4\sum_{i<j}\sum_{k<l} A_{ijkl}B_{ij}B_{kl}\right] \\ &= \frac{1}{4}\sum_{i=1}^{q_2}\sum_{j=1}^{q_2}\sum_{k=1}^{q_2}\sum_{l=1}^{q_2} A_{ijkl}B_{ij}B_{kl}, \end{aligned}$$

where the properties $A_{ijkl} = A_{jikl}$, $A_{ijkl} = A_{ijlk}$, and $B_{ij} = B_{ji}$ for all i,j,k,l, are used in the last equality.

Straightforward computations (for the third one, use property (47)) now show that

$$-\frac{1}{2}T^{1/2}(\tilde{\pi}-\tilde{\pi}^*)'\mathcal{J}_{T,\tilde{\pi}\tilde{\pi}}T^{1/2}(\tilde{\pi}-\tilde{\pi}^*)=\frac{1}{2}(\tilde{\pi}-\tilde{\pi}^*)'\nabla^2_{\tilde{\pi}\tilde{\pi}'}L_T^{\pi}(\alpha,\tilde{\pi}^*,0)(\tilde{\pi}-\tilde{\pi}^*),$$

$$\begin{aligned}
&-T^{1/2}(\tilde{\pi}-\tilde{\pi}^*)'\mathcal{J}'_{T,\tilde{\pi}\varpi\varpi}T^{1/2}\alpha_1\alpha_2 v(\varpi)\\
&\quad=(\tilde{\pi}-\tilde{\pi}^*)'\left[c_{11}\nabla^3_{\tilde{\pi}\varpi_1\varpi_1}L_T^{\pi}(\alpha,\tilde{\pi}^*,0)\varpi_1^2+...+c_{q_2-1,q_2}\nabla^3_{\tilde{\pi}\varpi_{q_2-1}\varpi_{q_2}}L_T^{\pi}(\alpha,\tilde{\pi}^*,0)\varpi_{q_2-1}\varpi_{q_2}\right]\\
&\quad=\frac{3}{3!}\sum_{i=1}^{q}\sum_{j=1}^{q_2}\sum_{k=1}^{q_2}\nabla^3_{\tilde{\pi}_i\varpi_j\varpi_k}L_T^{\pi}(\alpha,\tilde{\pi}^*,0)(\tilde{\pi}_i-\tilde{\pi}_i^*)\varpi_j\varpi_k,
\end{aligned}$$

$$-\frac{1}{2}T\alpha_1^2\alpha_2^2 v(\varpi)'\mathcal{J}_{T,\varpi\varpi\varpi\varpi}v(\varpi)=\frac{1}{4!}\sum_{i=1}^{q_2}\sum_{j=1}^{q_2}\sum_{k=1}^{q_2}\sum_{l=1}^{q_2}\nabla^4_{\varpi_i\varpi_j\varpi_k\varpi_l}L_T^{\pi}(\alpha,\tilde{\pi}^*,0)\varpi_i\varpi_j\varpi_k\varpi_l.$$

Therefore the fourth-order Taylor expansion of $L_T^{\pi}(\alpha,\tilde{\pi},\varpi)$ in (36) can be written as a quadratic expansion given by

$$L_T^{\pi}(\alpha,\tilde{\pi},\varpi)-L_T^{\pi}(\alpha,\tilde{\pi}^*,0)=S_T'\boldsymbol{\theta}(\alpha,\tilde{\pi},\varpi)-\frac{1}{2}[T^{1/2}\boldsymbol{\theta}(\alpha,\tilde{\pi},\varpi)]'\mathcal{J}_T[T^{1/2}\boldsymbol{\theta}(\alpha,\tilde{\pi},\varpi)]+R_T^{(1)}(\alpha,\tilde{\pi},\varpi). \quad (48)$$

Next, define

$$\mathcal{I}=\begin{bmatrix}\mathcal{I}_{\tilde{\pi}\tilde{\pi}} & \mathcal{I}'_{\tilde{\pi}\varpi\varpi}\\ \mathcal{I}_{\tilde{\pi}\varpi\varpi} & \mathcal{I}_{\varpi\varpi\varpi\varpi}\end{bmatrix}, \quad (49)$$

where the matrices $\mathcal{I}_{\tilde{\pi}\tilde{\pi}}$ $((q_1+q_2)\times(q_1+q_2))$, $\mathcal{I}'_{\tilde{\pi}\varpi\varpi}$ $((q_1+q_2)\times q_\vartheta)$, and $\mathcal{I}_{\varpi\varpi\varpi\varpi}$ $(q_\vartheta\times q_\vartheta)$ are defined as follows

$$\begin{aligned}
\mathcal{I}_{\tilde{\pi}\tilde{\pi}}&=E\left[\frac{\nabla f_t(\tilde{\pi}^*)}{f_t(\tilde{\pi}^*)}\frac{\nabla' f_t(\tilde{\pi}^*)}{f_t(\tilde{\pi}^*)}\right],\\
\mathcal{I}'_{\tilde{\pi}\varpi\varpi}&=\left[c_{ij}E\left[\frac{\nabla f_t(\tilde{\pi}^*)}{f_t(\tilde{\pi}^*)}X^*_{t,i,j}\right]\right]_{(i,j)\in\mathfrak{I}},\\
\mathcal{I}_{\varpi\varpi\varpi\varpi}&=\left[c_{ij}c_{kl}E\left[X^*_{t,i,j}X^*_{t,k,l}\right]\right]_{(i,j,k,l)\in\mathfrak{I}\times\mathfrak{I}},
\end{aligned}$$

and where we have used the short-hand notation $X^*_{t,i,j}$, $i,j\in\{1,\dots,q_2\}$ (see (42)). Finiteness of $\mathcal{I}$ follows from Lemma F.1. Now, defining

$$R_T^{(2)}(\alpha,\tilde{\pi},\varpi)=-\frac{1}{2}[T^{1/2}\boldsymbol{\theta}(\alpha,\tilde{\pi},\varpi)]'(\mathcal{J}_T-\mathcal{I})[T^{1/2}\boldsymbol{\theta}(\alpha,\tilde{\pi},\varpi)]$$

and adding and subtracting terms, expansion (48) can be written as (37) with

$$R_T(\alpha,\tilde{\pi},\varpi)=R_T^{(1)}(\alpha,\tilde{\pi},\varpi)+R_T^{(2)}(\alpha,\tilde{\pi},\varpi).$$

F.3 Some more explicit derivatives and the verification of Assumption 5(iii)

Some more explicit derivative expressions. We will require more explicit expressions for the components of s_t in (41) (see also (38)–(40) and (42)). Straightforward computation shows that (as before, $\nabla f_t(\cdot)$ denotes differentiation of $f_t(\cdot)$ in (3) with respect to $\tilde{\phi}=(\tilde{\phi}_0,\tilde{\phi}_1,\dots,\tilde{\phi}_p,\tilde{\sigma}_1^2)$)

$$\frac{\nabla f_t(\tilde{\phi})}{f_t(\tilde{\phi})}=\nabla\log(f_t(\tilde{\phi})),\qquad \frac{\nabla^2 f_t(\tilde{\phi})}{f_t(\tilde{\phi})}=\nabla^2\log(f_t(\tilde{\phi}))+\nabla\log(f_t(\tilde{\phi}))\nabla'\log(f_t(\tilde{\phi})),$$

where (as $\log(f_t(\tilde{\phi})) = -\frac{1}{2}\log(2\bar{\pi}) - \frac{1}{2}\log(\tilde{\sigma}_1^2) - \frac{1}{2}g_t^2(\tilde{\phi})$ with $g_t(\tilde{\phi}) = [y_t - (\tilde{\phi}_0 + \tilde{\phi}_1 y_{t-1} + \cdots + \tilde{\phi}_p y_{t-p})]/\tilde{\sigma}_1$)

$$\nabla \log(f_t(\tilde{\phi})) = \begin{bmatrix} \frac{1}{\tilde{\sigma}_1} g_t(\tilde{\phi}) \\ \frac{1}{\tilde{\sigma}_1} \boldsymbol{y}_{t-1} g_t(\tilde{\phi}) \\ \frac{1}{2\tilde{\sigma}_1^2}(g_t^2(\tilde{\phi}) - 1) \end{bmatrix}, \qquad \nabla^2 \log(f_t(\tilde{\phi})) = \begin{bmatrix} -\frac{1}{\tilde{\sigma}_1^2} & -\frac{1}{\tilde{\sigma}_1^2}\boldsymbol{y}_{t-1}' & -\frac{1}{\tilde{\sigma}_1^3} g_t(\tilde{\phi}) \\ -\frac{1}{\tilde{\sigma}_1^2}\boldsymbol{y}_{t-1} & -\frac{1}{\tilde{\sigma}_1^2}\boldsymbol{y}_{t-1}\boldsymbol{y}_{t-1}' & -\frac{1}{\tilde{\sigma}_1^3} g_t(\tilde{\phi})\boldsymbol{y}_{t-1} \\ -\frac{1}{\tilde{\sigma}_1^3} g_t(\tilde{\phi}) & -\frac{1}{\tilde{\sigma}_1^3} g_t(\tilde{\phi})\boldsymbol{y}_{t-1}' & -\frac{1}{2\tilde{\sigma}_1^4}(2g_t^2(\tilde{\phi}) - 1) \end{bmatrix},$$

so that

$$\frac{\nabla f_t(\pi^*)}{f_t(\pi^*)} = \begin{bmatrix} \frac{1}{\tilde{\sigma}_1^*}\varepsilon_t \\ \frac{1}{\tilde{\sigma}_1^*}\boldsymbol{y}_{t-1}\varepsilon_t \\ \frac{1}{2\tilde{\sigma}_1^{*2}}(\varepsilon_t^2 - 1) \end{bmatrix},$$

$$\begin{aligned} \frac{\nabla^2 f_t(\pi^*)}{f_t(\pi^*)} &= \begin{bmatrix} -\frac{1}{\tilde{\sigma}_1^{*2}} & -\frac{1}{\tilde{\sigma}_1^{*2}}\boldsymbol{y}_{t-1}' & -\frac{1}{\tilde{\sigma}_1^{*3}}\varepsilon_t \\ -\frac{1}{\tilde{\sigma}_1^{*2}}\boldsymbol{y}_{t-1} & -\frac{1}{\tilde{\sigma}_1^{*2}}\boldsymbol{y}_{t-1}\boldsymbol{y}_{t-1}' & -\frac{1}{\tilde{\sigma}_1^{*3}}\boldsymbol{y}_{t-1}\varepsilon_t \\ -\frac{1}{\tilde{\sigma}_1^{*3}}\varepsilon_t & -\frac{1}{\tilde{\sigma}_1^{*3}}\boldsymbol{y}_{t-1}'\varepsilon_t & -\frac{1}{2\tilde{\sigma}_1^{*4}}(2\varepsilon_t^2 - 1) \end{bmatrix} + \begin{bmatrix} \frac{1}{\tilde{\sigma}_1^*}\varepsilon_t \\ \frac{1}{\tilde{\sigma}_1^*}\boldsymbol{y}_{t-1}\varepsilon_t \\ \frac{1}{2\tilde{\sigma}_1^{*2}}(\varepsilon_t^2 - 1) \end{bmatrix} \begin{bmatrix} \frac{1}{\tilde{\sigma}_1^*}\varepsilon_t \\ \frac{1}{\tilde{\sigma}_1^*}\boldsymbol{y}_{t-1}\varepsilon_t \\ \frac{1}{2\tilde{\sigma}_1^{*2}}(\varepsilon_t^2 - 1) \end{bmatrix}' \\ &= \begin{bmatrix} \frac{1}{\tilde{\sigma}_1^{*2}}(\varepsilon_t^2 - 1) & \frac{1}{\tilde{\sigma}_1^{*2}}\boldsymbol{y}_{t-1}'(\varepsilon_t^2 - 1) & \frac{1}{2\tilde{\sigma}_1^{*3}}(\varepsilon_t^3 - 3\varepsilon_t) \\ \frac{1}{\tilde{\sigma}_1^{*2}}\boldsymbol{y}_{t-1}(\varepsilon_t^2 - 1) & \frac{1}{\tilde{\sigma}_1^{*2}}\boldsymbol{y}_{t-1}\boldsymbol{y}_{t-1}'(\varepsilon_t^2 - 1) & \frac{1}{2\tilde{\sigma}_1^{*3}}\boldsymbol{y}_{t-1}(\varepsilon_t^3 - 3\varepsilon_t) \\ \frac{1}{2\tilde{\sigma}_1^{*3}}(\varepsilon_t^3 - 3\varepsilon_t) & \frac{1}{2\tilde{\sigma}_1^{*3}}\boldsymbol{y}_{t-1}'(\varepsilon_t^3 - 3\varepsilon_t) & \frac{1}{4\tilde{\sigma}_1^{*4}}(\varepsilon_t^4 - 6\varepsilon_t^2 + 3) \end{bmatrix}. \end{aligned} \tag{50}$$

Similar formulas hold for $\mathsf{n}_p(\boldsymbol{y}_{t-1};\tilde{\phi})$. As $\log(\mathsf{n}_p(\boldsymbol{y}_{t-1};\tilde{\phi})) = -\frac{p}{2}\log(2\bar{\pi}) - \frac{1}{2}\log(\det(\boldsymbol{\Gamma}_{1,p})) - \frac{1}{2}(\boldsymbol{y}_{t-1} - \mu_1\mathbf{1}_p)'\boldsymbol{\Gamma}_{1,p}^{-1}(\boldsymbol{y}_{t-1} - \mu_1\mathbf{1}_p)$ (see Section 2.2), we obtain, for each $i = 1,\ldots,p+2$, (cf. Magnus and Neudecker (1999, p. 325))[13]

$$\nabla_i \log(\mathsf{n}_p(\boldsymbol{y}_{t-1};\tilde{\phi})) = \frac{1}{2}tr\left(\frac{\partial\boldsymbol{\Gamma}_{1,p}^{-1}}{\partial\tilde{\phi}_i}\boldsymbol{\Gamma}_{1,p}\right) + (\boldsymbol{y}_{t-1} - \mu_1\mathbf{1}_p)'\boldsymbol{\Gamma}_{1,p}^{-1}(\frac{\partial\mu_1}{\partial\tilde{\phi}_i}\mathbf{1}_p) - \frac{1}{2}(\boldsymbol{y}_{t-1} - \mu_1\mathbf{1}_p)'\frac{\partial\boldsymbol{\Gamma}_{1,p}^{-1}}{\partial\tilde{\phi}_i}(\boldsymbol{y}_{t-1} - \mu_1\mathbf{1}_p),$$

where (see Section 2.2 for the notation)

$$\frac{\partial\mu_1}{\partial\tilde{\phi}} = \begin{bmatrix} (\tilde{\phi}(1))^{-1} \\ \tilde{\phi}_0(\tilde{\phi}(1))^{-2}\mathbf{1}_p \\ 0 \end{bmatrix}.$$

For the expression of $\frac{\partial\boldsymbol{\Gamma}_{1,p}^{-1}}{\partial\tilde{\phi}_i}$, first note that $\boldsymbol{\Gamma}_{1,p}^{-1}$ can be expressed as (see, e.g., Galbraith and Galbraith (1974)) $\boldsymbol{\Gamma}_{1,p}^{-1} = \frac{1}{\tilde{\sigma}_1^2}(U'U - V'V)$ with U and V being $p \times p$ Toeplitz matrices given by

$$U = \begin{bmatrix} 1 & & & \\ -\tilde{\phi}_1 & \ddots & & \\ \vdots & \ddots & \ddots & \\ -\tilde{\phi}_{p-1} & \cdots & -\tilde{\phi}_1 & 1 \end{bmatrix}, \quad V = \begin{bmatrix} \tilde{\phi}_p & & & \\ \tilde{\phi}_{p-1} & \ddots & & \\ \vdots & \ddots & \ddots & \\ \tilde{\phi}_1 & \cdots & \tilde{\phi}_{p-1} & \tilde{\phi}_p \end{bmatrix}.$$

Thus $\frac{\partial\boldsymbol{\Gamma}_{1,p}^{-1}}{\partial\tilde{\phi}_i}$ equals a zero matrix when differentiating with respect to $\tilde{\phi}_0$, $-\frac{1}{\tilde{\sigma}_1^2}\boldsymbol{\Gamma}_{1,p}^{-1}$ when differentiating

[13] As before, ∇ denotes differentiation with respect to $\tilde{\phi} = (\tilde{\phi}_0, \tilde{\phi}_1, \ldots, \tilde{\phi}_p, \tilde{\sigma}_1^2)$ and ∇_i, $i = 1,\ldots,p+2$, with respect to the ith component of $\tilde{\phi}$.

with respect to $\tilde{\sigma}_1^2$, and

$$\frac{\partial \boldsymbol{\Gamma}_{1,p}^{-1}}{\partial \tilde{\phi}_i} = \frac{1}{\tilde{\sigma}_1^2}\left(\frac{\partial U'}{\partial \tilde{\phi}_i}U + U'\frac{\partial U}{\partial \tilde{\phi}_i} - \frac{\partial V'}{\partial \tilde{\phi}_i}V - V'\frac{\partial V}{\partial \tilde{\phi}_i}\right)$$

when differentiating with respect to to the autoregressive parameters. To summarize,

$$\frac{\nabla \mathsf{n}_p(\tilde{\phi})}{\mathsf{n}_p(\tilde{\phi})} = \nabla \log(\mathsf{n}_p(\boldsymbol{y}_{t-1};\tilde{\phi})) = \begin{bmatrix} d_1(\boldsymbol{y}_{t-1};\tilde{\phi}) \\ d_2(\boldsymbol{y}_{t-1};\tilde{\phi}) \\ d_3(\boldsymbol{y}_{t-1};\tilde{\phi}) \end{bmatrix}, \tag{51}$$

where (note that $tr\left(\frac{\partial \boldsymbol{\Gamma}_{1,p}^{-1}}{\partial \tilde{\phi}_i}\boldsymbol{\Gamma}_{1,p}\right) = \frac{\partial vec(\boldsymbol{\Gamma}_{1,p}^{-1})'}{\partial \tilde{\phi}_i} vec(\boldsymbol{\Gamma}_{1,p})$)

$$\begin{aligned}
d_1(\boldsymbol{y}_{t-1};\tilde{\phi}) &= (\tilde{\phi}(1))^{-1}(\boldsymbol{y}_{t-1} - \mu_1\mathbf{1}_p)'\boldsymbol{\Gamma}_{1,p}^{-1}\mathbf{1}_p, \\
d_2(\boldsymbol{y}_{t-1};\tilde{\phi}) &= \frac{1}{2}\frac{\partial vec(\boldsymbol{\Gamma}_{1,p}^{-1})'}{\partial(\tilde{\phi}_1,\ldots,\tilde{\phi}_p)} vec(\boldsymbol{\Gamma}_{1,p}) + \tilde{\phi}_0(\tilde{\phi}(1))^{-2}\mathbf{1}_p(\mathbf{1}_p'\boldsymbol{\Gamma}_{1,p}^{-1}(\boldsymbol{y}_{t-1} - \mu_1\mathbf{1}_p)) \\
&\quad - \frac{1}{2}\frac{\partial vec(\boldsymbol{\Gamma}_{1,p}^{-1})'}{\partial(\tilde{\phi}_1,\ldots,\tilde{\phi}_p)}((\boldsymbol{y}_{t-1} - \mu_1\mathbf{1}_p)\otimes(\boldsymbol{y}_{t-1} - \mu_1\mathbf{1}_p)), \\
d_3(\boldsymbol{y}_{t-1};\tilde{\phi}) &= -\frac{p}{2\tilde{\sigma}_1^2} + \frac{1}{2\tilde{\sigma}_1^2}(\boldsymbol{y}_{t-1} - \mu_1\mathbf{1}_p)'\boldsymbol{\Gamma}_{1,p}^{-1}(\boldsymbol{y}_{t-1} - \mu_1\mathbf{1}_p)
\end{aligned}$$

(first and last scalars, middle one $p \times 1$). Therefore

$$\frac{\nabla \mathsf{n}_p(\pi^*)}{\mathsf{n}_p(\pi^*)} = \nabla \log(\mathsf{n}_p(\boldsymbol{y}_{t-1};\pi^*)) = \begin{bmatrix} d_1(\boldsymbol{y}_{t-1};\pi^*) \\ d_2(\boldsymbol{y}_{t-1};\pi^*) \\ d_3(\boldsymbol{y}_{t-1};\pi^*) \end{bmatrix}.$$

Based on the preceding derivations, the derivatives appearing in (38)–(40) can now be expressed as

$$\begin{aligned}
\nabla_\beta l_t^\pi(\alpha,\beta^*,\pi^*,0) &= \frac{1}{\tilde{\sigma}_1^*}\varepsilon_t \\
\nabla_\pi l_t^\pi(\alpha,\beta^*,\pi^*,0) &= \begin{bmatrix} \frac{1}{\tilde{\sigma}_1^*}\boldsymbol{y}_{t-1}\varepsilon_t \\ \frac{1}{2\tilde{\sigma}_1^{*2}}(\varepsilon_t^2 - 1) \end{bmatrix} \qquad ((p+1)\times 1) \\
\nabla^2_{\varpi\varpi'} l_t^\pi(\alpha,\beta^*,\pi^*,0) &= \alpha(1-\alpha)\Bigg\{ \begin{bmatrix} d_2(\boldsymbol{y}_{t-1};\pi^*)\frac{1}{\tilde{\sigma}_1^*}\boldsymbol{y}_{t-1}'\varepsilon_t & d_2(\boldsymbol{y}_{t-1};\pi^*)\frac{1}{2\tilde{\sigma}_1^{*2}}(\varepsilon_t^2 - 1) \\ d_3(\boldsymbol{y}_{t-1};\pi^*)\frac{1}{\tilde{\sigma}_1^*}\boldsymbol{y}_{t-1}'\varepsilon_t & d_3(\boldsymbol{y}_{t-1};\pi^*)\frac{1}{2\tilde{\sigma}_1^{*2}}(\varepsilon_t^2 - 1) \end{bmatrix} \\
&\quad + \begin{bmatrix} \frac{1}{\tilde{\sigma}_1^*}\boldsymbol{y}_{t-1}\varepsilon_t d_2'(\boldsymbol{y}_{t-1};\pi^*) & \frac{1}{\tilde{\sigma}_1^*}\boldsymbol{y}_{t-1}\varepsilon_t d_3(\boldsymbol{y}_{t-1};\pi^*) \\ \frac{1}{2\tilde{\sigma}_1^{*2}}(\varepsilon_t^2 - 1)d_2'(\boldsymbol{y}_{t-1};\pi^*) & \frac{1}{2\tilde{\sigma}_1^{*2}}(\varepsilon_t^2 - 1)d_3(\boldsymbol{y}_{t-1};\pi^*) \end{bmatrix} \\
&\quad + \begin{bmatrix} \frac{1}{\tilde{\sigma}_1^{*2}}\boldsymbol{y}_{t-1}\boldsymbol{y}_{t-1}'(\varepsilon_t^2 - 1) & \frac{1}{2\tilde{\sigma}_1^{*3}}\boldsymbol{y}_{t-1}(\varepsilon_t^3 - 3\varepsilon_t) \\ \frac{1}{2\tilde{\sigma}_1^{*3}}\boldsymbol{y}_{t-1}'(\varepsilon_t^3 - 3\varepsilon_t) & \frac{1}{4\tilde{\sigma}_1^{*4}}(\varepsilon_t^4 - 6\varepsilon_t^2 + 3) \end{bmatrix}\Bigg\}.
\end{aligned}$$

Verification of Assumption 5(iii). Finiteness of $\mathcal{I}$ was already established in Supplementary Appendix F.2. For positive definiteness, it suffices to show that the components of the vector s_t are linearly independent. Note that for linear independence, it does not matter if the order of the elements is changed or if some of the elements are multiplied by nonzero constants. Therefore, making use of the explicit expressions given above, it suffices to show that the components of the vector $\tilde{s}_t = (\tilde{s}_{t,1}, \tilde{s}_{t,2}, \tilde{s}_{t,3}, \tilde{s}_{t,4}, \tilde{s}_{t,5}, \tilde{s}_{t,6})$ (where the dimensions of the six components are $1, p, 1, p(p+1)/2, p, 1$, respectively) are linearly independent, where

$$\begin{bmatrix} \tilde{s}_{t,1} \\ \tilde{s}_{t,2} \\ \tilde{s}_{t,3} \\ \tilde{s}_{t,4} \\ \tilde{s}_{t,5} \\ \tilde{s}_{t,6} \end{bmatrix} = \begin{bmatrix} \frac{1}{\tilde{\sigma}_1^*}\varepsilon_t \\ \frac{1}{\tilde{\sigma}_1^*}\boldsymbol{y}_{t-1}\varepsilon_t \\ \frac{1}{2\tilde{\sigma}_1^{*2}}(\varepsilon_t^2 - 1) \\ vech[d_2(\boldsymbol{y}_{t-1};\pi^*)\boldsymbol{y}_{t-1}' + \boldsymbol{y}_{t-1}d_2'(\boldsymbol{y}_{t-1};\pi^*)]\frac{1}{\tilde{\sigma}_1^*}\varepsilon_t + vech[\boldsymbol{y}_{t-1}\boldsymbol{y}_{t-1}']\frac{1}{\tilde{\sigma}_1^{*2}}(\varepsilon_t^2 - 1) \\ d_2(\boldsymbol{y}_{t-1};\pi^*)\frac{1}{2\tilde{\sigma}_1^{*2}}(\varepsilon_t^2 - 1) + \boldsymbol{y}_{t-1}d_3(\boldsymbol{y}_{t-1};\pi^*)\frac{1}{\tilde{\sigma}_1^*}\varepsilon_t + \frac{1}{2\tilde{\sigma}_1^{*3}}\boldsymbol{y}_{t-1}(\varepsilon_t^3 - 3\varepsilon_t) \\ d_3(\boldsymbol{y}_{t-1};\pi^*)\frac{1}{\tilde{\sigma}_1^{*2}}(\varepsilon_t^2 - 1) + \frac{1}{4\tilde{\sigma}_1^{*4}}(\varepsilon_t^4 - 6\varepsilon_t^2 + 3) \end{bmatrix}.$$

To this end, suppose that $c'\tilde{s}_t = (c_1, c_2, c_3, c_4, c_5, c_6)'(\tilde{s}_{t,1}, \tilde{s}_{t,2}, \tilde{s}_{t,3}, \tilde{s}_{t,4}, \tilde{s}_{t,5}, \tilde{s}_{t,6}) = 0$ (with the dimension of c and its subvectors chosen conformably). Note that the only random quantities $\tilde{s}_t$ depends on are $\boldsymbol{y}_{t-1}$ and ε_t which are independent. First, as the term ε_t^4 only appears in $\tilde{s}_{t,6}$, the equality $E[c'\tilde{s}_t \mid \varepsilon_t] = 0$ can be expressed as $c_6\varepsilon_t^4/(4\tilde{\sigma}_1^{*4}) + P_3(\varepsilon_t) = 0$, where $P_3(\varepsilon_t)$ is a third-order polynomial in ε_t. As the components of the vector $(\varepsilon_t, \varepsilon_t^2, \varepsilon_t^3, \varepsilon_t^4)$ are linearly independent (this clearly follows from normality), it follows that $c_6 = 0$. Next, basic properties of the standard normal distribution imply that $E[c'\tilde{s}_t(\varepsilon_t^3 - 3\varepsilon_t) \mid \boldsymbol{y}_{t-1}] = c_5'\frac{1}{2\tilde{\sigma}_1^{*3}}\boldsymbol{y}_{t-1}E[(\varepsilon_t^3 - 3\varepsilon_t)^2] = 0$, so that necessarily $c_5 = 0$ (as the components of $\boldsymbol{y}_{t-1}$ are linearly independent and $E[(\varepsilon_t^3 - 3\varepsilon_t)^2] > 0$). Next note that (as $c_5 = 0$, $c_6 = 0$)

$$0 = E[c'\tilde{s}_t(\varepsilon_t^2 - 1) \mid \boldsymbol{y}_{t-1}] = c_3E[(\varepsilon_t^2 - 1)^2]/(2\tilde{\sigma}_1^{*2}) + c_4'vech[\boldsymbol{y}_{t-1}\boldsymbol{y}_{t-1}']E[(\varepsilon_t^2 - 1)^2]/\tilde{\sigma}_1^{*2}$$

so that $c_4'vech[\boldsymbol{y}_{t-1}\boldsymbol{y}_{t-1}'] = -c_3/2$. As the components of $vech[\boldsymbol{y}_{t-1}\boldsymbol{y}_{t-1}']$ are linearly independent (as $vech(\boldsymbol{y}_{t-1}\boldsymbol{y}_{t-1}') = D_p^+vec(\boldsymbol{y}_{t-1}\boldsymbol{y}_{t-1}') = D_p^+(\boldsymbol{y}_{t-1} \otimes \boldsymbol{y}_{t-1})$, with D_p^+ denoting the Moore-Penrose inverse of the duplication matrix D_p, $Cov[vech(\boldsymbol{y}_{t-1}\boldsymbol{y}_{t-1}')] = D_p^+Cov(\boldsymbol{y}_{t-1} \otimes \boldsymbol{y}_{t-1})D_p^{+\prime}$; because D_p^+ is of full row rank and $Cov(\boldsymbol{y}_{t-1} \otimes \boldsymbol{y}_{t-1})$ has rank $p(p+1)/2$ (Magnus and Neudecker (1979, Thm 4.3(v)), $Cov[vech(\boldsymbol{y}_{t-1}\boldsymbol{y}_{t-1}')]$ is positive definite), it necessarily follows that $c_4 = 0$ and $c_3 = 0$. Finally, as only c_1 and c_2 may be nonzero, $E[c'\tilde{s}_t\varepsilon_t \mid \boldsymbol{y}_{t-1}] = c_1\frac{1}{\tilde{\sigma}_1^*} + c_2'\frac{1}{\tilde{\sigma}_1^*}\boldsymbol{y}_{t-1} = 0$, from which $c_2 = 0$ and $c_1 = 0$ follow (as the components of $\boldsymbol{y}_{t-1}$ are linearly independent).

F.4 Verification of Assumption 5(iv)

First consider $R_T^{(1)}(\alpha, \tilde{\pi}, \varpi)$. Of the quantities on the right hand side of (46), the first two are equal to zero because $\nabla_{\varpi}l_t^{\pi}(\alpha, \tilde{\pi}^*, 0) = 0$ and $\nabla_{\tilde{\pi}\varpi'}l_t^{\pi}(\alpha, \tilde{\pi}^*, 0) = 0$; for the other eight quantities, Lemma F.4 provides upper bounds that aid in bounding them. Now, to verify Assumption 5(iv) (for $R_T^{(1)}(\alpha, \tilde{\pi}, \varpi)$), let $\{\gamma_T,\ T \geq 1\}$ be an arbitrary sequence of (non-random) positive scalars such that $\gamma_T \to 0$ as $T \to \infty$. Condition $\|(\beta, \pi, \varpi) - (\beta^*, \pi^*, 0)\| \leq \gamma_T$ (appearing in Assumption 5(iv)), together with the properties $\|\boldsymbol{\theta}(\alpha, \beta, \pi, \varpi)\|^2 = \|\tilde{\pi} - \tilde{\pi}^*\|^2 + \alpha_1^2\alpha_2^2\|v(\varpi)\|^2$ and $\|v(\varpi)\| \leq \|\varpi\|^2$, implies that

$$\|\boldsymbol{\theta}(\alpha, \beta, \pi, \varpi)\|^{1/2} \leq \|\tilde{\pi} - \tilde{\pi}^*\|^{1/2} + \alpha_1^{1/2}\alpha_2^{1/2}\|v(\varpi)\|^{1/2} \leq \gamma_T^{1/2} + \alpha_1^{1/2}\alpha_2^{1/2}\gamma_T.$$

This, Lemma F.4, and the fact that α is bounded away from zero and one on A, imply that for some sequence $\{\tilde{\gamma}_T,\ T \geq 1\}$ of (non-random) positive scalars such that $\tilde{\gamma}_T \to 0$ as $T \to \infty$ and for some finite C,

$$\begin{aligned}
&\sup_{\alpha\in A}\sup_{\substack{(\tilde{\pi},\varpi)\in B\times\Pi_\alpha,\\ \|(\tilde{\pi},\varpi)-(\tilde{\pi}^*,0)\|\leq\gamma_T}}\frac{|R_T^{(1)}(\alpha,\tilde{\pi},\varpi)|}{\left(1+\left\|T^{1/2}\boldsymbol{\theta}(\alpha,\tilde{\pi},\varpi)\right\|\right)^2}\\
&\leq\tilde{\gamma}_T\sum_{i,j,k}\sup_{\alpha\in A}\left[|T^{-1}\nabla^3_{\tilde{\pi}_i\tilde{\pi}_j\tilde{\pi}_k}L_T^\pi(\alpha,\tilde{\pi}^*,0)|+|T^{-1}\nabla^3_{\tilde{\pi}_i\tilde{\pi}_j\varpi_k}L_T^\pi(\alpha,\tilde{\pi}^*,0)|+|T^{-1/2}\nabla^3_{\varpi_i\varpi_j\varpi_k}L_T^\pi(\alpha,\tilde{\pi}^*,0)|\right]\\
&+\tilde{\gamma}_T\sum_{i,j,k,l}\sup_{\alpha\in A}\sup_{\substack{(\tilde{\pi},\varpi)\in B\times\Pi_\alpha,\\ \|(\tilde{\pi},\varpi)-(\tilde{\pi}^*,0)\|\leq\gamma_T}}\left[|T^{-1}\nabla^4_{\tilde{\pi}_i\tilde{\pi}_j\tilde{\pi}_k\tilde{\pi}_l}L_T^\pi(\alpha,\tilde{\pi},\varpi)|+|T^{-1}\nabla^4_{\tilde{\pi}_i\tilde{\pi}_j\tilde{\pi}_k\varpi_l}L_T^\pi(\alpha,\tilde{\pi},\varpi)|\right.\\
&\qquad\qquad\left.+|T^{-1}\nabla^4_{\tilde{\pi}_i\tilde{\pi}_j\varpi_k\varpi_l}L_T^\pi(\alpha,\tilde{\pi},\varpi)|+|T^{-1}\nabla^4_{\tilde{\pi}_i\varpi_j\varpi_k\varpi_l}L_T^\pi(\alpha,\tilde{\pi},\varpi)|\right]\\
&+C\sum_{i,j,k,l}\sup_{\alpha\in A}\sup_{\substack{(\tilde{\pi},\varpi)\in B\times\Pi_\alpha,\\ \|(\tilde{\pi},\varpi)-(\tilde{\pi}^*,0)\|\leq\gamma_T}}|T^{-1}(\nabla^4_{\varpi_i\varpi_j\varpi_k\varpi_l}L_T^\pi(\alpha,\tilde{\pi},\varpi)-\nabla^4_{\varpi_i\varpi_j\varpi_k\varpi_l}L_T^\pi(\alpha,\tilde{\pi}^*,0))|, \qquad (52)
\end{aligned}$$

where the summations above are understood to contain counterparts of each term in (46). As the data is assumed to be generated by a linear autoregression (Assumption 1), the y_t's form a stationary and ergodic process. Moreover, as the reparameterized log-likelihood of the GMAR model is four times continuously differentiable (see Assumption 4 and its verification), also the $\nabla^3_{\tilde{\pi}_i\tilde{\pi}_j\tilde{\pi}_k}l_t^\pi(\alpha,\tilde{\pi}^*,0)$'s form a stationary and ergodic process (for any i,j,k). An analogous result holds for all the third and fourth partial derivatives of $l_t^\pi(\alpha,\tilde{\pi},\varpi)$ appearing on the majorant side of (52).

Now, Lemma F.2(iii) together with the ergodic theorem implies that $\sup_{\alpha\in A}[|T^{-1}\nabla^3_{\tilde{\pi}_i\tilde{\pi}_j\tilde{\pi}_k}L_T^\pi(\alpha,\tilde{\pi}^*,0)|]=O_p(1)$ (for any i,j,k). Similarly, $\sup_{\alpha\in A}[|T^{-1}\nabla^3_{\tilde{\pi}_i\tilde{\pi}_j\varpi_k}L_T^\pi(\alpha,\tilde{\pi}^*,0)|]=O_p(1)$ (for any i,j,k). Expression of $\nabla^3_{\varpi_i\varpi_j\varpi_k}l_t^\pi(\alpha,\tilde{\pi}^*,0)$ in Supplementary Appendix F.1, Lemmas F.1 and F.3, and the compactness of A, imply that $\sup_{\alpha\in A}|T^{-1/2}\nabla^3_{\varpi_i\varpi_j\varpi_k}L_T^\pi(\alpha,\tilde{\pi}^*,0)|\leq C|T^{-1/2}\sum_{t=1}^T MDS_{t,i,j,k}(\pi^*)|$ for some finite C and for some square integrable martingale difference sequence $MDS_{t,i,j,k}(\pi^*)$. Moreover, for any i,j,k, the last upper bound is $O_p(1)$ by an appropriate central limit theorem (Billingsley (1961)).

As for the fourth-order partial derivatives appearing on the majorant side of (52), Lemma F.2(iv) and a uniform law of large numbers for stationary and ergodic processes (Ranga Rao (1962)) imply that $\sup_{\alpha\in A}\sup_{(\tilde{\pi},\varpi)\in B\times\Pi_\alpha,\|(\tilde{\pi},\varpi)-(\tilde{\pi}^*,0)\|\leq\gamma_T}|T^{-1}\nabla^4_{\tilde{\pi}_i\tilde{\pi}_j\tilde{\pi}_k\tilde{\pi}_l}L_T^\pi(\alpha,\tilde{\pi},\varpi)|=O_p(1)$ (for any i,j,k,l). The next three terms in (52) can be handled similarly. As for the last term on the majorant side of (52),

$$\begin{aligned}
&\sup_{\alpha\in A}\sup_{\substack{(\tilde{\pi},\varpi)\in B\times\Pi_\alpha,\\ \|(\tilde{\pi},\varpi)-(\tilde{\pi}^*,0)\|\leq\gamma_T}}|T^{-1}(\nabla^4_{\varpi_i\varpi_j\varpi_k\varpi_l}L_T^\pi(\alpha,\tilde{\pi},\varpi)-\nabla^4_{\varpi_i\varpi_j\varpi_k\varpi_l}L_T^\pi(\alpha,\tilde{\pi}^*,0))|\\
&\qquad\leq 2\sup_{\alpha\in A}\sup_{\substack{(\tilde{\pi},\varpi)\in B\times\Pi_\alpha,\\ \|(\tilde{\pi},\varpi)-(\tilde{\pi}^*,0)\|\leq\gamma_T}}|T^{-1}\nabla^4_{\varpi_i\varpi_j\varpi_k\varpi_l}L_T^\pi(\alpha,\tilde{\pi},\varpi)-E[\nabla^4_{\varpi_i\varpi_j\varpi_k\varpi_l}l_t^\pi(\alpha,\tilde{\pi},\varpi)]|
\end{aligned}$$

where the dominant side is $o_p(1)$ (again relying on Lemma F.2(iv) and a uniform LLN). To conclude, the upper bound in (52) is $\tilde{\gamma}_T O_p(1)+Co_p(1)=o_p(1)$. This completes the verification of Assumption 5(iv) for the term $R_T^{(1)}(\alpha,\tilde{\pi},\varpi)$.

Now consider $R_T^{(2)}(\alpha,\tilde{\pi},\varpi)=-\frac{1}{2}[T^{1/2}\boldsymbol{\theta}(\alpha,\tilde{\pi},\varpi)]'(\mathcal{J}_T-\mathcal{I})[T^{1/2}\boldsymbol{\theta}(\alpha,\tilde{\pi},\varpi)]$. We will below show that (a) $\mathcal{J}_T\xrightarrow{p}\mathcal{J}$ as $T\to\infty$, where the matrix $\mathcal{J}$ will be specified below (and $\mathcal{J}_T,\mathcal{J}$ do not depend on α). Write (-2 times) $R_T^{(2)}(\alpha,\tilde{\pi},\varpi)$ as

$$[T^{1/2}\boldsymbol{\theta}(\alpha,\tilde{\pi},\varpi)]'(\mathcal{J}_T-\mathcal{J})[T^{1/2}\boldsymbol{\theta}(\alpha,\tilde{\pi},\varpi)]+[T^{1/2}\boldsymbol{\theta}(\alpha,\tilde{\pi},\varpi)]'(\mathcal{J}-\mathcal{I})[T^{1/2}\boldsymbol{\theta}(\alpha,\tilde{\pi},\varpi)].$$

We will below also show that (b) the latter term above equals zero. The validity of Assumption 5(iv) for the term $R_T^{(2)}(\alpha,\tilde{\pi},\varpi)$ follows from results (a) and (b) (together with usual properties of the Euclidean norm).

To prove claim (a), we first define the matrix $\mathcal{J}$ as

$$\mathcal{J} = \begin{bmatrix} \mathcal{J}_{\pi\pi} & \mathcal{J}'_{\pi\varpi\varpi} \\ \mathcal{J}_{\pi\varpi\varpi} & \mathcal{J}_{\varpi\varpi\varpi\varpi} \end{bmatrix}$$

where the matrices $\mathcal{J}_{\tilde{\pi}\tilde{\pi}}$ $((q_1+q_2)\times(q_1+q_2))$, $\mathcal{J}'_{\tilde{\pi}\varpi\varpi}$ $((q_1+q_2)\times q_\vartheta)$, and $\mathcal{J}_{\varpi\varpi\varpi\varpi}$ $(q_\vartheta \times q_\vartheta)$ are defined as

$$\begin{aligned}
\mathcal{J}_{\tilde{\pi}\tilde{\pi}} &= E\left[\frac{\nabla f_t^*}{f_t^*}\frac{\nabla' f_t^*}{f_t^*}\right] \\
\mathcal{J}'_{\tilde{\pi}\varpi\varpi} &= E\left[\left[c_{ij}\frac{\nabla f_t^*}{f_t^*}X^*_{t,i,j}\right]_{(i,j)\in\mathfrak{I}}\right] \\
\mathcal{J}_{\varpi\varpi\varpi\varpi} &= \frac{1}{3}\left[c_{ij}c_{kl}\left(E[X^*_{t,i,j}X^*_{t,k,l}] + E[X^*_{t,i,k}X^*_{t,j,l}] + E[X^*_{t,i,l}X^*_{t,j,k}]\right)\right]_{(i,j,k,l)\in\mathfrak{I}\times\mathfrak{I}}
\end{aligned}$$

where the $X^*_{t,i,j}$ $(i,j\in\{1,\ldots,q_2\})$ are as in (42). Finiteness of $\mathcal{J}$ follows from Lemma F.1.

Now consider the convergence result $\mathcal{J}_T \xrightarrow{p} \mathcal{J}$ for each block at a time. For the top-left block, from Supplementary Appendix F.1 we have $\nabla_{\tilde{\pi}\tilde{\pi}'}l_t^{\pi*} = \frac{\nabla^2 f_t^*}{f_t^*} - \frac{\nabla f_t^*}{f_t^*}\frac{\nabla' f_t^*}{f_t^*}$ so that ergodic theorem and Lemmas F.1 and F.3 (latter ensuring the first term on the right-hand side of the previous equation has zero expectation) imply that $\mathcal{J}_{T,\tilde{\pi}\tilde{\pi}} = -T^{-1}\nabla^2_{\tilde{\pi}\tilde{\pi}'}L_T^{\pi*} \xrightarrow{p} \mathcal{J}_{\tilde{\pi}\tilde{\pi}}$.

For the off-diagonal block, consider the expression of $\nabla^3_{\tilde{\pi}_i\varpi_j\varpi_k}l_t^{\pi*}$ in Supplementary Appendix F.1. Lemma F.3 ensures that of the ten summands in this expression, only the second, fourth, and sixth ones have non-zero expectation. Therefore the ergodic theorem and Lemma F.1 imply that

$$\mathcal{J}'_{T,\tilde{\pi}\varpi\varpi} = -T^{-1}\frac{1}{\alpha_1\alpha_2}[c_{11}\nabla^3_{\tilde{\pi}\varpi_1\varpi_1}L_T^{\pi*} : \cdots : c_{q_2-1,q_2}\nabla^3_{\tilde{\pi}\varpi_{q_2-1}\varpi_{q_2}}L_T^{\pi*}] \xrightarrow{p} \mathcal{J}'_{\tilde{\pi}\varpi\varpi}.$$

Lastly, for the bottom-right block, consider the expression of $\nabla^4_{\varpi_i\varpi_j\varpi_k\varpi_l}l_t^{\pi*}$ in Supplementary Appendix F.1. Lemma F.3 reveals that the terms in this expression that have non-zero expectation can be expressed as

$$-\alpha_1^2\alpha_2^2[X^*_{t,i,j}X^*_{t,k,l} + X^*_{t,i,k}X^*_{t,j,l} + X^*_{t,i,l}X^*_{t,j,k}].$$

Therefore the ergodic theorem and Lemma F.1 imply that

$$\mathcal{J}_{T,\varpi\varpi\varpi\varpi} = -T^{-1}\frac{8}{4!}\frac{1}{\alpha_1^2\alpha_2^2}\left[c_{ij}c_{kl}\nabla^4_{\varpi_i\varpi_j\varpi_k\varpi_l}L_T^{\pi*}\right]_{(i,j,k,l)\in\mathfrak{I}\times\mathfrak{I}} \xrightarrow{p} \mathcal{J}_{\varpi\varpi\varpi\varpi}.$$

This completes the proof of claim (a).

To prove claim (b), first note that from the definitions of $\mathcal{J}$ and $\mathcal{I}$ (see (49)) it can be seen that only the bottom-right blocks of $\mathcal{J}$ and $\mathcal{I}$ differ. Therefore, as $T^{1/2}\boldsymbol{\theta}(\alpha,\tilde{\pi},\varpi) = (T^{1/2}(\tilde{\pi}-\tilde{\pi}^*), T^{1/2}(\alpha_1\alpha_2 v(\varpi)))$, claim (b) holds if $T(\alpha_1\alpha_2)^2 v(\varpi)'(\mathcal{J}_{\varpi\varpi\varpi\varpi} - \mathcal{I}_{\varpi\varpi\varpi\varpi})v(\varpi) = 0$ where

$$\begin{aligned}
\mathcal{J}_{\varpi\varpi\varpi\varpi} &= \frac{1}{3}\left[c_{ij}c_{kl}\left(E[X^*_{t,i,j}X^*_{t,k,l}] + E[X^*_{t,i,k}X^*_{t,j,l}] + E[X^*_{t,i,l}X^*_{t,j,k}]\right)\right]_{(i,j,k,l)\in\mathfrak{I}\times\mathfrak{I}}, \\
\mathcal{I}_{\varpi\varpi\varpi\varpi} &= \left[c_{ij}c_{kl}E[X^*_{t,i,j}X^*_{t,k,l}]\right]_{(i,j,k,l)\in\mathfrak{I}\times\mathfrak{I}}.
\end{aligned}$$

Note that the scalars $A_{ijkl} = E[X^*_{t,i,j}X^*_{t,k,l}]$ satisfy $A_{ijkl} = A_{jikl}$ and $A_{ijkl} = A_{ijlk}$ for all i,j,k,l so that using property (47) we obtain

$$\begin{aligned}
v(\varpi)'\mathcal{J}_{\varpi\varpi\varpi\varpi}v(\varpi) &= \frac{1}{3}v(\varpi)'\left[c_{ij}c_{kl}\left(E[X^*_{t,i,j}X^*_{t,k,l}] + E[X^*_{t,i,k}X^*_{t,j,l}] + E[X^*_{t,i,l}X^*_{t,j,k}]\right)\right]_{(i,j,k,l)\in\mathfrak{I}\times\mathfrak{I}}v(\varpi) \\
&= \frac{1}{3}\frac{1}{4}\sum_{i=1}^{q_2}\sum_{j=1}^{q_2}\sum_{k=1}^{q_2}\sum_{l=1}^{q_2}\left(E[X^*_{t,i,j}X^*_{t,k,l}] + E[X^*_{t,i,k}X^*_{t,j,l}] + E[X^*_{t,i,l}X^*_{t,j,k}]\right)\varpi_i\varpi_j\varpi_k\varpi_l \\
&= \frac{1}{4}\sum_{i=1}^{q_2}\sum_{j=1}^{q_2}\sum_{k=1}^{q_2}\sum_{l=1}^{q_2}E[X^*_{t,i,j}X^*_{t,k,l}]\varpi_i\varpi_j\varpi_k\varpi_l \\
&= v(\varpi)'\mathcal{I}_{\varpi\varpi\varpi\varpi}v(\varpi).
\end{aligned}$$

This completes the proof of claim (b).

Therefore, the verification of Assumption 5(iv) is done.

F.5 Additional Lemmas

The following four lemmas contain results needed in the proofs. Note that the first and the third lemma are not specific to the examples in this paper, whereas the second and fourth lemmas concern only the GMAR example. In the first lemma, $\mathsf{n}_{p+1}(\tilde{\phi}) = \mathsf{n}_{p+1}(y_t, \boldsymbol{y}_{t-1}; \tilde{\phi})$ denotes the $(p+1)$-dimensional density function of an AR(p) process based on parameter value $\tilde{\phi}$ evaluated at $(y_t, \boldsymbol{y}_{t-1})$; cf. equations (4)–(6) for the p-dimensional counterpart $\mathsf{n}_p(\tilde{\phi}) = \mathsf{n}_p(\boldsymbol{y}_{t-1}; \tilde{\phi})$.

Lemma F.1. *For any $i, j, k, l \in \{1, \ldots, p+2\}$ and any positive r, the following moments are all finite:*

(i) $E\big[\sup_{\tilde{\phi}\in\tilde{\Phi}}|\nabla_i f_t(\tilde{\phi})/f_t(\tilde{\phi})|^r\big]$, $E\big[\sup_{\tilde{\phi}\in\tilde{\Phi}}|\nabla^2_{ij} f_t(\tilde{\phi})/f_t(\tilde{\phi})|^r\big], \ldots,$ $E\big[\sup_{\tilde{\phi}\in\tilde{\Phi}}|\nabla^4_{ijkl} f_t(\tilde{\phi})/f_t(\tilde{\phi})|^r\big]$,

(ii) $E\big[\sup_{\tilde{\phi}\in\tilde{\Phi}}|\nabla_i \mathsf{n}_p(\tilde{\phi})/\mathsf{n}_p(\tilde{\phi})|^r\big]$, $E\big[\sup_{\tilde{\phi}\in\tilde{\Phi}}|\nabla^2_{ij} \mathsf{n}_p(\tilde{\phi})/\mathsf{n}_p(\tilde{\phi})|^r\big], \ldots,$ $E\big[\sup_{\tilde{\phi}\in\tilde{\Phi}}|\nabla^4_{ijkl} \mathsf{n}_p(\tilde{\phi})/\mathsf{n}_p(\tilde{\phi})|^r\big]$,

(iii) $E\big[\sup_{\tilde{\phi}\in\tilde{\Phi}}|\nabla_i \mathsf{n}_{p+1}(\tilde{\phi})|^r\big]$, $E\big[\sup_{\tilde{\phi}\in\tilde{\Phi}}|\nabla^2_{ij} \mathsf{n}_{p+1}(\tilde{\phi})|^r\big], \ldots,$ $E\big[\sup_{\tilde{\phi}\in\tilde{\Phi}}|\nabla^4_{ijkl} \mathsf{n}_{p+1}(\tilde{\phi})|^r\big]$.

Lemma F.2. *In the GMAR example the following hold, where each of (the scalars) z_1, z_2, z_3, z_4 is a 'placeholder' for any of $\tilde{\pi}_i, \tilde{\pi}_j, \tilde{\pi}_k, \tilde{\pi}_l$ $(i, j, k, l \in \{1, \ldots, q_1 + q_2\})$ or $\varpi_i, \varpi_j, \varpi_k, \varpi_l$ $(i, j, k, l \in \{1, \ldots, q_2\})$:*

(i) $E\left[\sup_{\alpha\in A}\sup_{(\tilde{\pi},\varpi)\in B\times\Pi_\alpha}|\nabla_{z_1} l^\pi_t(\alpha, \tilde{\pi}, \varpi)|\right] < \infty$,

(ii) $E\left[\sup_{\alpha\in A}\sup_{(\tilde{\pi},\varpi)\in B\times\Pi_\alpha}|\nabla^2_{z_1z_2} l^\pi_t(\alpha, \tilde{\pi}, \varpi)|\right] < \infty$,

(iii) $E\left[\sup_{\alpha\in A}\sup_{(\tilde{\pi},\varpi)\in B\times\Pi_\alpha}|\nabla^3_{z_1z_2z_3} l^\pi_t(\alpha, \tilde{\pi}, \varpi)|\right] < \infty$,

(iv) $E\left[\sup_{\alpha\in A}\sup_{(\tilde{\pi},\varpi)\in B\times\Pi_\alpha}|\nabla^4_{z_1z_2z_3z_4} l^\pi_t(\alpha, \tilde{\pi}, \varpi)|\right] < \infty$.

Lemma F.3. *For any $i, j, k, l \in \{1, \ldots, p+2\}$,*

$$E\Big[\ \frac{\nabla_i f^*_t}{f^*_t}\ \mid \boldsymbol{y}_{t-1}\Big] = E\Big[\ \frac{\nabla^2_{ij} f^*_t}{f^*_t}\ \mid \boldsymbol{y}_{t-1}\Big] = E\Big[\ \frac{\nabla^3_{ijk} f^*_t}{f^*_t}\ \mid \boldsymbol{y}_{t-1}\Big] = E\Big[\ \frac{\nabla^4_{ijkl} f^*_t}{f^*_t}\ \mid \boldsymbol{y}_{t-1}\Big] = 0.$$

Lemma F.4. *In the GMAR example the following hold for all* $\alpha \in A$, $(\tilde{\pi}, \varpi) \in B \times \Pi_\alpha$, T, *and* $i, j, k, l \in \{1, \ldots, q_1 + q_2\}$ *(subindex in* $\tilde{\pi}$*) or* $i, j, k, l \in \{1, \ldots, q_2\}$ *(subindex in* ϖ*):*

$$(i) \quad T\,|\tilde{\pi}_i - \tilde{\pi}_i^*|\,|\tilde{\pi}_j - \tilde{\pi}_j^*||\tilde{\pi}_k - \tilde{\pi}_k^*| \leq (1 + \|T^{1/2}\boldsymbol{\theta}(\alpha, \tilde{\pi}, \varpi)\|)^2 \,\|\tilde{\pi} - \tilde{\pi}^*\|\,,$$

$$(ii) \quad T|\tilde{\pi}_i - \tilde{\pi}_i^*||\tilde{\pi}_j - \tilde{\pi}_j^*||\varpi_k| \leq (1 + \|T^{1/2}\boldsymbol{\theta}(\alpha, \tilde{\pi}, \varpi)\|)^2 \,\|\varpi\|\,,$$

$$(iii) \quad T^{1/2}|\varpi_i||\varpi_j||\varpi_k| \leq (1 + \|T^{1/2}\boldsymbol{\theta}(\alpha, \tilde{\pi}, \varpi)\|)^2 (\alpha_1\alpha_2)^{-3/2} \,\|\boldsymbol{\theta}(\alpha, \tilde{\pi}, \varpi)\|^{1/2}\,,$$

$$(iv) \quad T(\tilde{\pi}_i - \tilde{\pi}_i^*)(\tilde{\pi}_j - \tilde{\pi}_j^*)(\tilde{\pi}_k - \tilde{\pi}_k^*)(\tilde{\pi}_l - \tilde{\pi}_l^*) \leq (1 + \|T^{1/2}\boldsymbol{\theta}(\alpha, \tilde{\pi}, \varpi)\|)^2 \,\|\tilde{\pi} - \tilde{\pi}^*\|^2\,,$$

$$(v) \quad T(\tilde{\pi}_i - \tilde{\pi}_i^*)(\tilde{\pi}_j - \tilde{\pi}_j^*)(\tilde{\pi}_k - \tilde{\pi}_k^*)\varpi_l \leq (1 + \|T^{1/2}\boldsymbol{\theta}(\alpha, \tilde{\pi}, \varpi)\|)^2 \,\|\tilde{\pi} - \tilde{\pi}^*\|\,\|\varpi\|\,,$$

$$(vi) \quad T(\tilde{\pi}_i - \tilde{\pi}_i^*)(\tilde{\pi}_j - \tilde{\pi}_j^*)\varpi_k\varpi_l \leq (1 + \|T^{1/2}\boldsymbol{\theta}(\alpha, \tilde{\pi}, \varpi)\|)^2 \,\|\varpi\|^2\,,$$

$$(vii) \quad T(\tilde{\pi}_i - \tilde{\pi}_i^*)\varpi_j\varpi_k\varpi_l \leq (1 + \|T^{1/2}\boldsymbol{\theta}(\alpha, \tilde{\pi}, \varpi)\|)^2 (\alpha_1\alpha_2)^{-3/2} \,\|\boldsymbol{\theta}(\alpha, \tilde{\pi}, \varpi)\|^{1/2}\,,$$

$$(viii) \quad T\varpi_i\varpi_j\varpi_k\varpi_l \leq (1 + \|T^{1/2}\boldsymbol{\theta}(\alpha, \tilde{\pi}, \varpi)\|)^2 (\alpha_1\alpha_2)^{-2}.$$

F.6 Proofs of Lemmas F.1–F.4

Proof of Lemma F.1. Writing $g_t(\tilde{\phi}) = [y_t - (\tilde{\phi}_0 + \tilde{\phi}_1 y_{t-1} + \cdots + \tilde{\phi}_p y_{t-p})]/\tilde{\sigma}_1$ and recalling the definition of $f_t(\tilde{\phi})$ we can write $f_t(\tilde{\phi}) = \tilde{\sigma}_1^{-1}\mathfrak{n}(g_t(\tilde{\phi}))$ where $\mathfrak{n}(\cdot)$ denotes the density function of a standard normal random variable. Recall also that derivatives of $\mathfrak{n}(\cdot)$ can be expressed using (one version of) Hermite polynomials $H_n(x)$ as $\frac{d^n}{dx^n}\mathfrak{n}(x) = (-1)^n H_n(x)\mathfrak{n}(x)$. Using the chain rule for differentiation repeatedly, it can therefore be seen that each of the functions $\nabla_i f_t(\tilde{\phi})/f_t(\tilde{\phi})$, $\nabla^2_{ij} f_t(\tilde{\phi})/f_t(\tilde{\phi})$, $\nabla^3_{ijk} f_t(\tilde{\phi})/f_t(\tilde{\phi})$, and $\nabla^4_{ijkl} f_t(\tilde{\phi})/f_t(\tilde{\phi})$ can be expressed as a sum of terms each of which is a product involving Hermite polynomials $H_n(g_t(\tilde{\phi}))$ and powers of derivatives of $g_t(\tilde{\phi})$ (and functions of $\tilde{\phi}$). Thus, each of these functions is a polynomial in terms of $y_t, y_{t-1}, \ldots, y_{t-p}$. As the y_t's are generated by a stationary linear Gaussian AR(p) model, they possess moments of all orders, implying (together with the definition of $\tilde{\Phi}$, implying in particular that $\tilde{\sigma}_1$ is bounded away from zero on $\tilde{\Phi}$) the finiteness of the moments listed in part (i) of the lemma.

As for part (ii), note that $\mathsf{n}_p(\tilde{\phi})$ can be expressed as $\mathsf{n}_p(\tilde{\phi}) = g_1(\tilde{\phi})\mathfrak{n}(g_{2,t}(\tilde{\phi}))$ for some function $g_1(\tilde{\phi})$ not depending on the y_t's and $g_{2,t}(\tilde{\phi})$ the square root of a second-order polynomial in $y_{t-1}, \ldots, y_{t-p}$. Therefore the finiteness of the moments listed in the part (ii) follows using similar arguments as above (noting that the definition of $\tilde{\Phi}$ implies that the determinant of the covariance matrix appearing in $\mathsf{n}_p(\tilde{\phi})$ is bounded away from zero on $\tilde{\Phi}$).

Finally, for part (iii), similar arguments, together with the observation that $\mathsf{n}_{p+1}(\tilde{\phi})$ is bounded on $\tilde{\Phi}$, yield the desired result. ■

Proof of Lemma F.2. To prove (i), first consider the derivatives with respect to ϖ. From the formulas in Supplementary Appendix F.7 we obtain

$$\begin{aligned}\nabla_\varpi l_t^\pi(\alpha, \tilde{\pi}, \varpi) = {} & \alpha_{1,t}\alpha_{2,t}\left(\frac{D^{(1)\prime}_{\tilde{\phi},\varpi}\nabla\mathsf{n}_p(\tilde{\phi})}{\mathsf{n}_p(\tilde{\phi})} - \frac{D^{(1)\prime}_{\tilde{\varphi},\varpi}\nabla\mathsf{n}_p(\tilde{\varphi})}{\mathsf{n}_p(\tilde{\varphi})}\right)\frac{f_t(\tilde{\phi})}{f^\pi_{2,t}(\alpha, \tilde{\pi}, \varpi)} + \frac{D^{(1)\prime}_{\tilde{\phi},\varpi}\nabla f_t(\tilde{\phi})}{f^\pi_{2,t}(\alpha, \tilde{\pi}, \varpi)}\alpha_{1,t} \\ & - \alpha_{1,t}\alpha_{2,t}\left(\frac{D^{(1)\prime}_{\tilde{\phi},\varpi}\nabla\mathsf{n}_p(\tilde{\phi})}{\mathsf{n}_p(\tilde{\phi})} - \frac{D^{(1)\prime}_{\tilde{\varphi},\varpi}\nabla\mathsf{n}_p(\tilde{\varphi})}{\mathsf{n}_p(\tilde{\varphi})}\right)\frac{f_t(\tilde{\varphi})}{f^\pi_{2,t}(\alpha, \tilde{\pi}, \varpi)} + \frac{D^{(1)\prime}_{\tilde{\varphi},\varpi}\nabla f_t(\tilde{\varphi})}{f^\pi_{2,t}(\alpha, \tilde{\pi}, \varpi)}(1 - \alpha_{1,t})\end{aligned}$$

where $\tilde{\phi}$ and $\tilde{\varphi}$ are understood as functions of $(\alpha, \tilde{\pi}, \varpi)$ (i.e., $\tilde{\phi} = (\beta, \pi + \alpha_2\varpi)$ and $\tilde{\varphi} = (\beta, \pi - \alpha_1\varpi)$). Note that whenever $\alpha \in A$ and $(\tilde{\pi}, \varpi) \in B \times \Pi_\alpha$, $\tilde{\phi} \in \tilde{\Phi}$ and $\tilde{\varphi} \in \tilde{\Phi}$. Also note that over $\alpha \in A$ and

$(\tilde{\pi}, \varpi) \in B \times \Pi_\alpha$, the quantities

$$|\alpha_{1,t}|, \ |\alpha_{2,t}|, \ \|D^{(1)}_{\tilde{\phi},\varpi}\|, \ \|D^{(1)}_{\tilde{\varphi},\varpi}\|, \ |f_t(\tilde{\phi})/f^\pi_{2,t}(\alpha,\tilde{\pi},\varpi)|, \ |f_t(\tilde{\varphi})/f^\pi_{2,t}(\alpha,\tilde{\pi},\varpi)|$$

are all bounded by finite constants. Therefore $E\big[\sup_{\alpha\in A}\sup_{(\tilde{\pi},\varpi)\in N(\tilde{\pi}^*,0)}\|\nabla_\varpi l^\pi_t(\alpha,\tilde{\pi},\varpi)\|\big] < \infty$ as long as $E\big[\sup_{\tilde{\phi}\in\tilde{\Phi}}\|\nabla f_t(\tilde{\phi})/f_t(\tilde{\phi})\|\big] < \infty$ and $E\big[\sup_{\tilde{\phi}\in\tilde{\Phi}}\|\nabla \mathsf{n}_p(\tilde{\phi})/\mathsf{n}_p(\tilde{\phi})\|\big] < \infty$, which is ensured by Lemma F.1. The argument for $\nabla_{\tilde{\pi}} l^\pi_t(\alpha,\tilde{\pi},\varpi)$ is entirely similar and is omitted.

To prove (ii)–(iv), entirely similar arguments can be used. Tedious calculations (details omitted) show that the finiteness of the required moments is ensured by the finiteness of the moments in Lemma F.1(i) and (ii). ■

Proof of Lemma F.3. For the first two derivatives, the stated result follows directly from the expressions of $\nabla f^*_t/f^*_t$ and $\nabla^2 f^*_t/f^*_t$ in (50). The results for the third and fourth derivatives can be obtained with straightforward calculation. ■

Proof of Lemma F.4. First recall that $\left\|T^{1/2}\boldsymbol{\theta}(\alpha,\tilde{\pi},\varpi)\right\|^2 = T\left\|\tilde{\pi}-\tilde{\pi}^*\right\|^2 + T\alpha_1^2\alpha_2^2\left\|v(\varpi)\right\|^2$. (i) By an elementary inequality, $T\left|\tilde{\pi}_i-\tilde{\pi}^*_i\right|\left|\tilde{\pi}_j-\tilde{\pi}^*_j\right|\left|\tilde{\pi}_k-\tilde{\pi}^*_k\right| \le T\left\|\tilde{\pi}-\tilde{\pi}^*\right\|^3$ and therefore the result follows by adding nonnegative terms on the majorant side of this inequality. Parts (ii) and (iv)–(vi) are shown similarly. (vii) As $\boldsymbol{\theta}(\alpha,\tilde{\pi},\varpi) = (\tilde{\pi}-\tilde{\pi}^*, \alpha_1\alpha_2 v(\varpi))$, each of the terms $|\tilde{\pi}_i-\tilde{\pi}^*_i|$, $\alpha_1\alpha_2\varpi_j^2$, $\alpha_1\alpha_2\varpi_k^2$, and $\alpha_1\alpha_2\varpi_l^2$ are dominated by $\|\boldsymbol{\theta}(\alpha,\tilde{\pi},\varpi)\|$. Therefore

$$T|\tilde{\pi}_i-\tilde{\pi}^*_i||\varpi_j||\varpi_k||\varpi_l| \le T(\alpha_1\alpha_2)^{-3/2}\left\|\boldsymbol{\theta}(\alpha,\tilde{\pi},\varpi)\right\|^{5/2} \le (\alpha_1\alpha_2)^{-3/2}(1+\left\|T^{1/2}\boldsymbol{\theta}(\alpha,\tilde{\pi},\varpi)\right\|)^2\left\|\boldsymbol{\theta}(\alpha,\tilde{\pi},\varpi)\right\|^{1/2}$$

where the second inequality holds because nonnegative terms were added to the majorant side. (viii) Similarly as in the previous part,

$$T|\varpi_i||\varpi_j||\varpi_k||\varpi_l| \le (\alpha_1\alpha_2)^{-2}\left\|T^{1/2}\boldsymbol{\theta}(\alpha,\tilde{\pi},\varpi)\right\|^2 \le (\alpha_1\alpha_2)^{-2}(1+\left\|T^{1/2}\boldsymbol{\theta}(\alpha,\tilde{\pi},\varpi)\right\|)^2.$$

Finally, for (iii) we, similarly as above but scaling with $T^{1/2}$ instead of T, obtain

$$T^{1/2}|\varpi_i||\varpi_j||\varpi_k| \le T^{1/2}(\alpha_1\alpha_2)^{-3/2}\left\|\boldsymbol{\theta}(\alpha,\tilde{\pi},\varpi)\right\|^{3/2} \le (\alpha_1\alpha_2)^{-3/2}(1+\left\|T^{1/2}\boldsymbol{\theta}(\alpha,\tilde{\pi},\varpi\alpha)\right\|)^2\left\|\boldsymbol{\theta}(\alpha,\tilde{\pi},\varpi)\right\|^{1/2},$$

which completes the proof. ■

F.7 Partial derivatives of the reparameterized log-likelihood function (continued)

Note that $l^\pi_t(\alpha,\beta,\pi,\varpi) = \log[f^\pi_{2,t}(\alpha,\beta,\pi,\varpi)]$ with

$$\begin{aligned}
f^\pi_{2,t}(\alpha,\beta,\pi,\varpi) &= \alpha^\pi_{1,t}(\alpha,\beta,\pi,\varpi)f_t(\beta,\pi+\alpha_2\varpi) + (1-\alpha^\pi_{1,t}(\alpha,\beta,\pi,\varpi))f_t(\beta,\pi-\alpha_1\varpi),\\
\alpha^\pi_{1,t}(\alpha,\beta,\pi,\varpi) &= \alpha^G_{1,t}(\alpha,(\beta,\pi+\alpha_2\varpi),(\beta,\pi-\alpha_1\varpi)).
\end{aligned}$$

For the sake of brevity, but with slight abuse of notation, we will write these as

$$\begin{aligned}
f^\pi_{2,t}(\alpha,\beta,\pi,\varpi) &= \alpha^\pi_{1,t}(\alpha,\beta,\pi,\varpi)f_t(\tilde{\phi}) + (1-\alpha^\pi_{1,t}(\alpha,\beta,\pi,\varpi))f_t(\tilde{\varphi}),\\
\alpha^\pi_{1,t}(\alpha,\beta,\pi,\varpi) &= \frac{\alpha\mathsf{n}_p(\tilde{\phi})}{\alpha\mathsf{n}_p(\tilde{\phi}) + (1-\alpha)\mathsf{n}_p(\tilde{\varphi})},
\end{aligned}$$

where $\tilde{\phi}$ and $\tilde{\varphi}$ are understood as functions of $(\alpha, \beta, \pi, \varpi)$, that is, $\tilde{\phi} = (\beta, \pi + \alpha_2\varpi)$ and $\tilde{\varphi} = (\beta, \pi - \alpha_1\varpi)$.

The following notation will be helpful:

$$
\begin{aligned}
D^{(1)}_{\tilde{\phi},\tilde{\pi}} &= \frac{\partial(\beta, \pi + \alpha_2\varpi)}{\partial\tilde{\pi}'} = I_{1+q_2} \\
D^{(1)}_{\tilde{\phi},\varpi} &= \frac{\partial(\beta, \pi + \alpha_2\varpi)}{\partial\varpi'} = \begin{bmatrix} 0 \\ \alpha_2 I_{q_2} \end{bmatrix} \quad ((1+q_2)\times q_2) \\
D^{(1)}_{\tilde{\varphi},\tilde{\pi}} &= \frac{\partial(\beta, \pi - \alpha_1\varpi)}{\partial\tilde{\pi}'} = I_{1+q_2} \\
D^{(1)}_{\tilde{\varphi},\varpi} &= \frac{\partial(\beta, \pi - \alpha_1\varpi)}{\partial\varpi'} = \begin{bmatrix} 0 \\ -\alpha_1 I_{q_2} \end{bmatrix} \quad ((1+q_2)\times q_2)
\end{aligned}
$$

First-order partial derivatives. With straightforward differentation we obtain

$$
\begin{aligned}
\nabla_{\tilde{\pi}} l_t^{\pi}(\alpha, \tilde{\pi}, \varpi) &= \frac{\nabla_{\tilde{\pi}} f_{2,t}^{\pi}(\alpha, \tilde{\pi}, \varpi)}{f_{2,t}^{\pi}(\alpha, \tilde{\pi}, \varpi)} \\
\nabla_{\varpi} l_t^{\pi}(\alpha, \tilde{\pi}, \varpi) &= \frac{\nabla_{\varpi} f_{2,t}^{\pi}(\alpha, \tilde{\pi}, \varpi)}{f_{2,t}^{\pi}(\alpha, \tilde{\pi}, \varpi)}
\end{aligned}
$$

with

$$
\begin{aligned}
\nabla_{\tilde{\pi}} f_{2,t}^{\pi}(\alpha, \tilde{\pi}, \varpi) &= \nabla_{\tilde{\pi}}\alpha_{1,t} f_t(\tilde{\phi}) + \nabla f_t(\tilde{\phi})\alpha_{1,t} - \nabla_{\tilde{\pi}}\alpha_{1,t} f_t(\tilde{\varphi}) + \nabla f_t(\tilde{\varphi})(1-\alpha_{1,t}) \\
\nabla_{\varpi} f_{2,t}^{\pi}(\alpha, \tilde{\pi}, \varpi) &= \nabla_{\varpi}\alpha_{1,t} f_t(\tilde{\phi}) + D^{(1)\prime}_{\tilde{\phi},\varpi}\nabla f_t(\tilde{\phi})\alpha_{1,t} - \nabla_{\varpi}\alpha_{1,t} f_t(\tilde{\varphi}) + D^{(1)\prime}_{\tilde{\varphi},\varpi}\nabla f_t(\tilde{\varphi})(1-\alpha_{1,t})
\end{aligned}
$$

and

$$
\begin{aligned}
\nabla_{\tilde{\pi}}\alpha_{1,t} &= \frac{\alpha_1 \nabla \mathsf{n}_p(\tilde{\phi})}{\alpha_1\mathsf{n}_p(\tilde{\phi}) + \alpha_2\mathsf{n}_p(\tilde{\varphi})} - \alpha_{1,t}\frac{\alpha_1\nabla\mathsf{n}_p(\tilde{\phi}) + \alpha_2\nabla\mathsf{n}_p(\tilde{\varphi})}{\alpha_1\mathsf{n}_p(\tilde{\phi}) + \alpha_2\mathsf{n}_p(\tilde{\varphi})} \\
\nabla_{\varpi}\alpha_{1,t} &= \frac{\alpha_1 D^{(1)\prime}_{\tilde{\phi},\varpi}\nabla \mathsf{n}_p(\tilde{\phi})}{\alpha_1\mathsf{n}_p(\tilde{\phi}) + \alpha_2\mathsf{n}_p(\tilde{\varphi})} - \alpha_{1,t}\frac{\alpha_1 D^{(1)\prime}_{\tilde{\phi},\varpi}\nabla\mathsf{n}_p(\tilde{\phi}) + \alpha_2 D^{(1)\prime}_{\tilde{\varphi},\varpi}\nabla\mathsf{n}_p(\tilde{\varphi})}{\alpha_1\mathsf{n}_p(\tilde{\phi}) + \alpha_2\mathsf{n}_p(\tilde{\varphi})}
\end{aligned}
$$

where simplification leads to

$$
\begin{aligned}
\nabla_{\tilde{\pi}}\alpha_{1,t} &= \alpha_{1,t}\frac{\nabla\mathsf{n}_p(\tilde{\phi})}{\mathsf{n}_p(\tilde{\phi})} - \alpha_{1,t}\left(\alpha_{1,t}\frac{\nabla\mathsf{n}_p(\tilde{\phi})}{\mathsf{n}_p(\tilde{\phi})} + \alpha_{2,t}\frac{\nabla\mathsf{n}_p(\tilde{\varphi})}{\mathsf{n}_p(\tilde{\varphi})}\right) \\
&= \alpha_{1,t}\alpha_{2,t}\left(\frac{\nabla\mathsf{n}_p(\tilde{\phi})}{\mathsf{n}_p(\tilde{\phi})} - \frac{\nabla\mathsf{n}_p(\tilde{\varphi})}{\mathsf{n}_p(\tilde{\varphi})}\right) \\
\nabla_{\varpi}\alpha_{1,t} &= \alpha_{1,t}\frac{D^{(1)\prime}_{\tilde{\phi},\varpi}\nabla\mathsf{n}_p(\tilde{\phi})}{\mathsf{n}_p(\tilde{\phi})} - \alpha_{1,t}\left(\alpha_{1,t}\frac{D^{(1)\prime}_{\tilde{\phi},\varpi}\nabla\mathsf{n}_p(\tilde{\phi})}{\mathsf{n}_p(\tilde{\phi})} + \alpha_{2,t}\frac{D^{(1)\prime}_{\tilde{\varphi},\varpi}\nabla\mathsf{n}_p(\tilde{\varphi})}{\mathsf{n}_p(\tilde{\varphi})}\right) \\
&= \alpha_{1,t}\alpha_{2,t}\left(\frac{D^{(1)\prime}_{\tilde{\phi},\varpi}\nabla\mathsf{n}_p(\tilde{\phi})}{\mathsf{n}_p(\tilde{\phi})} - \frac{D^{(1)\prime}_{\tilde{\varphi},\varpi}\nabla\mathsf{n}_p(\tilde{\varphi})}{\mathsf{n}_p(\tilde{\varphi})}\right).
\end{aligned}
$$

Evaluated at $(\alpha, \tilde{\pi}, \varpi) = (\alpha, \tilde{\pi}^*, 0)$ we get

$$\nabla_{\tilde{\pi}} \alpha^*_{1,t} = 0, \qquad \nabla_{\varpi} \alpha^*_{1,t} = \alpha_1 \alpha_2 \frac{\nabla_{(2,\ldots,p+2)} \mathsf{n}_p(\tilde{\pi}^*)}{\mathsf{n}_p(\tilde{\pi}^*)},$$

$$f^{\pi}_{2,t}(\alpha, \tilde{\pi}^*, 0) = f_t(\tilde{\pi}^*), \qquad \nabla_{\tilde{\pi}} f^{\pi}_{2,t}(\alpha, \tilde{\pi}^*, 0) = \nabla f_t(\tilde{\pi}^*), \qquad \nabla_{\varpi} f^{\pi}_{2,t}(\alpha, \tilde{\pi}^*, 0) = 0,$$

so that

$$\nabla_{\tilde{\pi}} l^{\pi}_t(\alpha, \tilde{\pi}^*, 0) = \frac{\nabla f_t(\tilde{\pi}^*)}{f_t(\tilde{\pi}^*)}, \qquad \nabla_{\varpi} l^{\pi}_t(\alpha, \tilde{\pi}^*, 0) = 0.$$

Second-order partial derivatives With straightforward differentation we obtain

$$\begin{aligned}
\nabla^2_{\tilde{\pi}\tilde{\pi}'} l^{\pi}_t(\alpha, \tilde{\pi}, \varpi) &= \frac{\nabla^2_{\tilde{\pi}\tilde{\pi}'} f^{\pi}_{2,t}(\alpha, \tilde{\pi}, \varpi)}{f^{\pi}_{2,t}(\alpha, \tilde{\pi}, \varpi)} - \frac{\nabla_{\tilde{\pi}} f^{\pi}_{2,t}(\alpha, \tilde{\pi}, \varpi)}{f^{\pi}_{2,t}(\alpha, \tilde{\pi}, \varpi)} \frac{\nabla_{\tilde{\pi}'} f^{\pi}_{2,t}(\alpha, \tilde{\pi}, \varpi)}{f^{\pi}_{2,t}(\alpha, \tilde{\pi}, \varpi)} \\
\nabla^2_{\tilde{\pi}\varpi'} l^{\pi}_t(\alpha, \tilde{\pi}, \varpi) &= \frac{\nabla^2_{\tilde{\pi}\varpi'} f^{\pi}_{2,t}(\alpha, \tilde{\pi}, \varpi)}{f^{\pi}_{2,t}(\alpha, \tilde{\pi}, \varpi)} - \frac{\nabla_{\tilde{\pi}} f^{\pi}_{2,t}(\alpha, \tilde{\pi}, \varpi)}{f^{\pi}_{2,t}(\alpha, \tilde{\pi}, \varpi)} \frac{\nabla_{\varpi'} f^{\pi}_{2,t}(\alpha, \tilde{\pi}, \varpi)}{f^{\pi}_{2,t}(\alpha, \tilde{\pi}, \varpi)} \\
\nabla^2_{\varpi\varpi'} l^{\pi}_t(\alpha, \tilde{\pi}, \varpi) &= \frac{\nabla^2_{\varpi\varpi'} f^{\pi}_{2,t}(\alpha, \tilde{\pi}, \varpi)}{f^{\pi}_{2,t}(\alpha, \tilde{\pi}, \varpi)} - \frac{\nabla_{\varpi} f^{\pi}_{2,t}(\alpha, \tilde{\pi}, \varpi)}{f^{\pi}_{2,t}(\alpha, \tilde{\pi}, \varpi)} \frac{\nabla_{\varpi'} f^{\pi}_{2,t}(\alpha, \tilde{\pi}, \varpi)}{f^{\pi}_{2,t}(\alpha, \tilde{\pi}, \varpi)}
\end{aligned}$$

with

$$\begin{aligned}
\nabla^2_{\tilde{\pi}\tilde{\pi}'} f^{\pi}_{2,t}(\alpha, \tilde{\pi}, \varpi) &= \nabla^2_{\tilde{\pi}\tilde{\pi}'} \alpha_{1,t} f_t(\tilde{\phi}) + \nabla_{\tilde{\pi}} \alpha_{1,t} \nabla' f_t(\tilde{\phi}) \\
&\quad + \nabla f_t(\tilde{\phi}) \nabla_{\tilde{\pi}'} \alpha_{1,t} + \alpha_{1,t} \nabla^2 f_t(\tilde{\phi}) \\
&\quad - \nabla^2_{\tilde{\pi}\tilde{\pi}'} \alpha_{1,t} f_t(\tilde{\varphi}) - \nabla_{\tilde{\pi}} \alpha_{1,t} \nabla' f_t(\tilde{\varphi}) \\
&\quad - \nabla f_t(\tilde{\varphi}) \nabla_{\tilde{\pi}'} \alpha_{1,t} + (1 - \alpha_{1,t}) \nabla^2 f_t(\tilde{\varphi}) \\
\nabla^2_{\tilde{\pi}\varpi'} f^{\pi}_{2,t}(\alpha, \tilde{\pi}, \varpi) &= \nabla^2_{\tilde{\pi}\varpi'} \alpha_{1,t} f_t(\tilde{\phi}) + \nabla_{\tilde{\pi}} \alpha_{1,t} \nabla' f_t(\tilde{\phi}) D^{(1)}_{\tilde{\phi},\varpi} \\
&\quad + \nabla f_t(\tilde{\phi}) \nabla_{\varpi'} \alpha_{1,t} + \alpha_{1,t} \nabla^2 f_t(\tilde{\phi}) D^{(1)}_{\tilde{\phi},\varpi} \\
&\quad - \nabla^2_{\tilde{\pi}\varpi'} \alpha_{1,t} f_t(\tilde{\varphi}) - \nabla_{\tilde{\pi}} \alpha_{1,t} \nabla' f_t(\tilde{\varphi}) D^{(1)}_{\tilde{\varphi},\varpi} \\
&\quad - \nabla f_t(\tilde{\varphi}) \nabla_{\varpi'} \alpha_{1,t} + (1 - \alpha_{1,t}) \nabla^2 f_t(\tilde{\varphi}) D^{(1)}_{\tilde{\varphi},\varpi} \\
\nabla^2_{\varpi\varpi'} f^{\pi}_{2,t}(\alpha, \tilde{\pi}, \varpi) &= \nabla^2_{\varpi\varpi'} \alpha_{1,t} f_t(\tilde{\phi}) + \nabla_{\varpi} \alpha_{1,t} \nabla' f_t(\tilde{\phi}) D^{(1)}_{\tilde{\phi},\varpi} \\
&\quad + \alpha_{1,t} D^{(1)\prime}_{\tilde{\phi},\varpi} \nabla^2 f_t(\tilde{\phi}) D^{(1)}_{\tilde{\phi},\varpi} + D^{(1)\prime}_{\tilde{\phi},\varpi} \nabla f_t(\tilde{\phi}) \nabla_{\varpi'} \alpha_{1,t} \\
&\quad - \nabla^2_{\varpi\varpi'} \alpha_{1,t} f_t(\tilde{\varphi}) - \nabla_{\varpi} \alpha_{1,t} \nabla' f_t(\tilde{\varphi}) D^{(1)}_{\tilde{\varphi},\varpi} \\
&\quad + (1 - \alpha_{1,t}) D^{(1)\prime}_{\tilde{\varphi},\varpi} \nabla^2 f_t(\tilde{\varphi}) D^{(1)}_{\tilde{\varphi},\varpi} - D^{(1)\prime}_{\tilde{\varphi},\varpi} \nabla f_t(\tilde{\varphi}) \nabla_{\varpi'} \alpha_{1,t}
\end{aligned}$$

For brevity, we omit the expressions of $\nabla^2_{\tilde{\pi}\tilde{\pi}'} \alpha_{1,t}$, $\nabla^2_{\tilde{\pi}\varpi'} \alpha_{1,t}$, and $\nabla^2_{\varpi\varpi'} \alpha_{1,t}$. Evaluated at $(\alpha, \tilde{\pi}, \varpi) = (\alpha, \tilde{\pi}^*, 0)$ we get

$$\begin{aligned}
\nabla^2_{\tilde{\pi}\tilde{\pi}'} f^{\pi}_{2,t}(\alpha, \tilde{\pi}^*, 0) &= \nabla^2 f_t(\tilde{\pi}^*) \\
\nabla^2_{\tilde{\pi}\varpi'} f^{\pi}_{2,t}(\alpha, \tilde{\pi}^*, 0) &= 0 \\
\nabla^2_{\varpi\varpi'} f^{\pi}_{2,t}(\alpha, \tilde{\pi}^*, 0) &= \nabla_{\varpi} \alpha^*_{1,t} \nabla'_{(2,\ldots,p+2)} f_t(\tilde{\pi}^*) + \nabla_{(2,\ldots,p+2)} f_t(\tilde{\pi}^*) \nabla_{\varpi'} \alpha^*_{1,t} + \alpha_1 \alpha_2 \nabla^2_{(2,\ldots,p+2)(2,\ldots,p+2)} f_t(\tilde{\pi}^*) \\
&= \nabla_{\varpi} \alpha^*_{1,t} \nabla'_{(2,\ldots,p+2)} f_t(\tilde{\pi}^*) + \nabla_{(2,\ldots,p+2)} f_t(\tilde{\pi}^*) \nabla_{\varpi'} \alpha^*_{1,t} + \alpha_1 \alpha_2 \nabla^2_{(2,\ldots,p+2)(2,\ldots,p+2)} f_t(\tilde{\pi}^*)
\end{aligned}$$

so that

$$
\begin{aligned}
\nabla^2_{\tilde{\pi}\tilde{\pi}'} l_t^{\pi}(\alpha, \tilde{\pi}^*, 0) &= \frac{\nabla^2 f_t(\tilde{\pi}^*)}{f_t(\tilde{\pi}^*)} - \frac{\nabla f_t(\tilde{\pi}^*)}{f_t(\tilde{\pi}^*)} \frac{\nabla' f_t(\tilde{\pi}^*)}{f_t(\tilde{\pi}^*)} \\
\nabla^2_{\tilde{\pi}\varpi'} l_t^{\pi}(\alpha, \tilde{\pi}^*, 0) &= 0 \\
\nabla^2_{\varpi\varpi'} l_t^{\pi}(\alpha, \tilde{\pi}^*, 0) &= \frac{\nabla_{\varpi}\alpha^*_{1,t} \nabla'_{(2,\ldots,p+2)} f_t(\tilde{\pi}^*) + \nabla_{(2,\ldots,p+2)} f_t(\tilde{\pi}^*) \nabla_{\varpi'}\alpha^*_{1,t} + \alpha_1\alpha_2 \nabla^2_{(2,\ldots,p+2)(2,\ldots,p+2)} f_t(\tilde{\pi}^*)}{f_t(\tilde{\pi}^*)} \\
&= \alpha_1\alpha_2 \left[\frac{\nabla_{(2,\ldots,p+2)} \mathsf{n}_p(\tilde{\pi}^*)}{\mathsf{n}_p(\tilde{\pi}^*)} \frac{\nabla'_{(2,\ldots,p+2)} f_t(\tilde{\pi}^*)}{f_t(\tilde{\pi}^*)} + \frac{\nabla_{(2,\ldots,p+2)} f_t(\tilde{\pi}^*)}{f_t(\tilde{\pi}^*)} \frac{\nabla'_{(2,\ldots,p+2)} \mathsf{n}_p(\tilde{\pi}^*)}{\mathsf{n}_p(\tilde{\pi}^*)} \right. \\
&\qquad \left. + \frac{\nabla^2_{(2,\ldots,p+2)(2,\ldots,p+2)} f_t(\tilde{\pi}^*)}{f_t(\tilde{\pi}^*)} \right].
\end{aligned}
$$

Additional References

Aliprantis, C. D., Burkinshaw, O., 1998. Principles of Real Analysis. Academic Press.

Andrews, D. W., Ploberger, W., 1994. Optimal tests when a nuisance parameter is present only under the alternative. Econometrica 62, 1383–1414.

Galbraith, R., Galbraith, J., 1974. On the inverses of some patterned matrices arising in the theory of stationary time series. Journal of Applied Probability 11, 63–71.

Magnus, J. R., Neudecker, H., 1979. The commutation matrix: some properties and applications. Annals of Statistics 7, 381–394.

Magnus, J. R., Neudecker, H., 1999. Matrix Differential Calculus with Applications in Statistics and Econometrics, Revised Edition. Wiley.