Electronic Journal of Statistics
Vol. 13 (2019) 5151–5212
ISSN: 1935-7524
https://doi.org/10.1214/19-EJS1636

Quasi-maximum likelihood estimation for cointegrated continuous-time linear state space models observed at low frequencies

Vicky Fasen-Hartmann*

Institute of Stochastics
Englerstraße 2
76131 Karlsruhe
Germany
e-mail: vicky.fasen@kit.edu

Markus Scholz*

Allianz Lebensversichung-AG
Reinsburgstraße 19
70197 Stuttgart
Germany
e-mail: scholz.ma@gmail.com

Abstract: In this paper, we investigate quasi-maximum likelihood (QML) estimation for the parameters of a cointegrated solution of a continuous-time linear state space model observed at discrete time points. The class of cointegrated solutions of continuous-time linear state space models is equivalent to the class of cointegrated continuous-time ARMA (MCARMA) processes. As a start, some pseudo-innovations are constructed to be able to define a QML-function. Moreover, the parameter vector is divided appropriately in long-run and short-run parameters using a representation for cointegrated solutions of continuous-time linear state space models as a sum of a Lévy process plus a stationary solution of a linear state space model. Then, we establish the consistency of our estimator in three steps. First, we show the consistency for the QML estimator of the long-run parameters. In the next step, we calculate its consistency rate. Finally, we use these results to prove the consistency for the QML estimator of the short-run parameters. After all, we derive the limiting distributions of the estimators. The long-run parameters are asymptotically mixed normally distributed, whereas the short-run parameters are asymptotically normally distributed. The performance of the QML estimator is demonstrated by a simulation study.

MSC 2010 subject classifications: Primary 91G70, 62H12; secondary 62M10, 60F05.
Keywords and phrases: Cointegration, (super-)consistency, identifiability, Kalman filter, MCARMA process, pseudo-innovation, quasi-maximum likelihood estimation, state space model.

Received September 2018.

*Financial support by the Deutsche Forschungsgemeinschaft through the research grant FA 809/2-2 is gratefully acknowledged.

Contents

1. Introduction

This paper deals with quasi-maximum likelihood (QML) estimation for the parameters of a cointegrated solution of a continuous-time linear state space model. The source of randomness in our model is a Lévy process, i.e., an $\mathbb{R}^m$-valued stochastic process $L = (L(t))_{t\geq 0}$ with $L(0) = 0_m$ $\mathbb{P}$-a.s., stationary and independent increments, and càdlàg sample paths. A typical example of a Lévy process is a Brownian motion. More details on Lévy processes can be found, e.g., in the monograph of Sato [48]. For deterministic matrices $A \in \mathbb{R}^{N\times N}$, $B \in \mathbb{R}^{N\times m}$, $C \in \mathbb{R}^{d\times N}$ and an $\mathbb{R}^m$-valued Lévy process L, an $\mathbb{R}^d$-valued continuous-time *linear state space model* (A, B, C, L) is defined by the state and observation equation

$$\begin{aligned} \mathrm{d}X(t) &= AX(t)\mathrm{d}t + B\mathrm{d}L(t), \\ Y(t) &= CX(t). \end{aligned} \tag{1.1}$$

The state vector process $X = (X(t))_{t\geq 0}$ is an $\mathbb{R}^N$-valued process and the output process $Y = (Y(t))_{t\geq 0}$ is an $\mathbb{R}^d$-valued process. Since in this model the driving noise is a Lévy process the model allows flexible margins. In particular, the margins can be Gaussian if we use a Brownian motion as Lévy process.

The topic of this paper are *cointegrated* solutions Y of linear state space models. Cointegrated means that Y is non stationary but has stationary increments, and there exist linear combinations of Y which are stationary. The cointegration space is the space spanned by all vectors β so that $\beta^\mathsf{T} Y$ is stationary. Without any transformation of the state space model (1.1) it is impossible to see clearly if there exists a cointegrated solution, not to mention the form of the cointegration space. In the case of a minimal state-space model (see Bernstein [9] for a definition), the eigenvalues of A determine whether a solution Y may be stationary or cointegrated. If the eigenvalue 0 of A has the same geometric and algebraic multiplicity $0 < c < \min(d, m)$, and all other eigenvalues of A have negative real parts, then there exists a cointegrated solution Y. In that case Y has the form

$$Y(t) = C_1 Z + C_1 B_1 L(t) + Y_{st}(t), \tag{1.2}$$

where $B_1 \in \mathbb{R}^{c\times m}$ and $C_1 \in \mathbb{R}^{d\times c}$ have rank c (see Fasen-Hartmann and Scholz [18, Theorem 3.3]). The starting vector Z is a c-dimensional random vector. The process $Y_{st} = (Y_{st}(t))_{t\geq 0}$ is a stationary solution of the state space model

$$\begin{aligned} \mathrm{d}X_{st}(t) &= A_2 X(t)\mathrm{d}t + B_2\mathrm{d}L(t), \\ Y_{st}(t) &= C_2 X_{st}(t), \end{aligned} \tag{1.3}$$

driven by the Lévy process L with $A_2 \in \mathbb{R}^{(N-c)\times(N-c)}$, $B_2 \in \mathbb{R}^{(N-c)\times m}$ and $C_2 \in \mathbb{R}^{d\times(N-c)}$. The matrices A, A_1, A_2, B, B_1, B_2, C, C_1, C_2 and C_3 are related through an invertible transformation matrix $T \in \mathbb{R}^{N\times N}$ such that

$$TAT^{-1} = \begin{pmatrix} 0_{c\times c} & 0_{c\times(N-c)} \\ 0_{(N-c)\times c} & A_2 \end{pmatrix} =: A', \quad TB = \left({B_1}^\mathsf{T}, {B_2}^\mathsf{T}\right)^\mathsf{T} =: B' \quad \text{and}$$
$$CT^{-1} = \begin{pmatrix} C_1, & C_2 \end{pmatrix} =: C',$$

where ${B_i}^\mathsf{T}$ denotes the transpose of B_i $(i = 1, 2)$ and $0_{(N-c)\times c} \in \mathbb{R}^{(N-c)\times c}$ denotes a matrix with only zero components. The process Y in (1.2) is obviously cointegrated with cointegration space spanned by the orthogonal of C_1 if the covariance matrix $\mathrm{Cov}(L(1))$ is non-singular. The probabilistic properties of Y

are analyzed in detail in Fasen-Hartmann and Scholz [18] and lay the groundwork for the present paper. Remarkable is that Y is a solution of the state space model (A', B', C', L) as well.

The class of cointegrated solutions of linear state space models is huge. They are equal to the class of cointegrated multivariate continuous-time ARMA (MCARMA) processes (see Fasen-Hartmann and Scholz [18]). As the name suggests, MCARMA processes are the continuous-time versions of the popular and well-established ARMA processes in discrete-time. In finance and economics continuous-time models provide the basis for option pricing, asset allocation and term structure theory. The underlying observations of asset prices, exchange rates, and interest rates are often irregularly spaced, in particular, in the context of high frequency data. Consequently, one often works with continuous-time models which infer the implied dynamics and properties of the estimated model at different frequencies (see Chen et al. [17]). Fitting discrete-time models to such kind of data have the drawback that the model parameters are not time-invariant: If the sampling frequency changes, then the parameters of the discrete-time model change as well. The advantages of continuous-time modelling over discrete-time modelling in economics and finance are described in detail, i.a., in the distinguished papers of Bergstrom [7], Phillips [43], Chambers, McCrorie and Thornton [15] and in signal processing, systems and control they are described in Sinha and Rao [54]. In particular, MCARMA models are applied in diversified fields as signal processing, systems and control (see Garnier and Wang [22], Sinha and Rao [53]), high-frequency financial econometrics (see Todorov [58]) and financial mathematics (see Benth et al. [6], Andresen et al. [1]). Thornton and Chambers [16] use them as well for modelling sunspot data. Empirical relevance of non-stationary MCARMA processes in economics and in finance is shown, i.a., in Thornton and Chambers [16, 56, 57].

There is not much known about the statistical inference of cointegrated Lévy driven MCARMA models. In the context of non-stationary MCARMA processes most attention is paid to Gaussian MCAR(p) (multivariate continuous-time AR) processes: An algorithm to estimate the structural parameters in a Gaussian MCAR(p) model by maximum-likelihood started already by Harvey and Stock [26, 27, 28] and were further explored in the well-known paper of Bergstrom [8]. Zadrozny [60] investigates continuous-time Brownian motion driven ARMAX models allowing stocks and flows at different frequencies and higher order integration. These papers use the state space representation of the MCARMA process and Kalman filtering techniques to compute the Gaussian likelihood function. In a recent paper Thornton and Chambers [57] extend the results to MCARMA processes with mixed stock-flow data using an exact discrete-time ARMA representation of the low-frequency observed MCARMA process. However, all of the papers have in common on the one hand, that they do not analyze the asymptotic properties of the estimators. On the other hand, they are not able to estimate the cointegration space directly or rather relate their results to cointegrated models.

Besides, statistical inference and identification of continuously and discretely observed cointegrated Gaussian MCAR(1) processes, which are homogeneous

Gaussian diffusions, are considered in Kessler and Rahbek [32, 33]; Stockmarr and Jacobsen [55] and frequency domain estimators for cointegrated Gaussian MCAR(p) models are topic of Chambers and McCrorie [14]. There are only a few papers investigating non-Gaussian cointegrated MCARMA processes. For example, Fasen [20] treats a multiple regression model in continuous-time. There the stationary part is a multivariate Ornstein-Uhlenbeck process and the process is observed on an equidistant time-grid. The model in Fasen [21] is similar but the stationary part is an MCARMA process and the process is observed on a high-frequency time grid.

The aim of this paper is to investigate QML estimators for C_1, B_1 and the parameters of the stationary process Y_{st} from the discrete-time observations $Y(h), \ldots, Y(nh)$ where $h > 0$ is fixed. The parameters of C_1 are the long-run parameters, whereas the other parameters are the short-run parameters. Although there exist results on QML for discrete-time cointegrated processes they can unfortunately not directly be applied to the sampled process for the following reasons.

MCARMA processes sampled equidistantly belong to the class of ARMA processes (see Thornton and Chambers [57] and Chambers, McCrorie and Thornton [15]). But identification problems arise from employing the ARMA structure for the estimation of MCARMA parameters. That is until now an unsolved problem (see as well the overview article Chambers, McCrorie and Thornton [15]). Moreover, in this representation the innovations are only uncorrelated and not iid (independent and identically distributed). However, statistical inference for cointegrated ARMA models has been done only for an iid noise elsewise even a Gaussian white noise, see, e.g., the monographs of Johansen [30], Lütkepohl [35] and Reinsel [44], and cannot be used for estimation of Lévy driven MCARMA processes.

Another attempt is to use the representation of the sampled continuous-time state space model as discrete-time state space model (see Zadrozny [60]). That is what we do in this paper. Sampling Y with distance $h > 0$ results in $Y^{(h)} := (Y(kh))_{k\in\mathbb{N}_0} =: (Y_k^{(h)})_{k\in\mathbb{N}_0}$, a cointegrated solution of the discrete time state-space model

$$\begin{aligned} X_k^{(h)} &= \mathrm{e}^{Ah} X_k^{(h)} + \xi_k^{(h)}, \\ Y_k^{(h)} &= C X_k^{(h)}, \end{aligned} \tag{1.4}$$

where $(\xi_k^{(h)})_{k\in\mathbb{N}_0} := (\int_{(k-1)h}^{kh} \mathrm{e}^{A(kh-t)} B \mathrm{d}L(t))_{k\in\mathbb{N}_0}$ is an iid sequence. For cointegrated solutions of discrete-time state space models of the form

$$\begin{aligned} X_k &= A X_k + \epsilon_k, \\ Y_k &= C X_k + \epsilon_k, \end{aligned} \tag{1.5}$$

where $(\epsilon_k)_{k\in\mathbb{N}_0}$ is a white noise, asymptotic properties of the QML estimator were investigated in the unpublished work of Bauer and Wagner [4]. An essential difference between the state space model (1.4) and (1.5) is that in (1.4) the noise is only going into the state equation, whereas in (1.5) it is going into both the state and the observation equation. An advantage of model (1.5) over our state space model is that it is already in innovation form, i.e., the white noise $(\epsilon_k)_{k\in\mathbb{N}_0}$ can be represented by finitely many past values of $(Y_k)_{k\in\mathbb{N}_0}$ due to

$$\epsilon_k = Y_k - C(A - BC)^k X_0 - C\sum_{j=1}^{k}(A - BC)^{j-1}BY_{k-j}. \tag{1.6}$$

But in our model (1.4) it is not possible to write the noise $(\xi_k^{(h)})_{k\in\mathbb{N}_0}$ by the past of $(Y_k^{(h)})_{k\in\mathbb{N}_0}$. Therefore, we are not able to apply the asymptotic results of Bauer and Wagner [4] to the setting of our paper.

We use the Kalman-filter to calculate the linear innovations and to construct an error correction form (see Fasen-Hartmann and Scholz [18, Proposition 5.5 and Theorem 5.8]). However, the linear innovations and the error correction form use infinitely many past values in contrast to the usual finite order form for VARMA models and discrete-time state space models as, e.g., in Lütkepohl and Claessen [36], Saikkonen [45], Yap and Reinsel [59] and respectively, Aoki [2], Bauer and Wagner [4] (see (1.6)). Indeed, the linear innovations are stationary, but in general it is not possible to say anything about their mixing properties. Hence, standard limit results for stationary mixing processes cannot be applied. For more details in the case of stationary MCARMA models we refer to Schlemm and Stelzer [50].

The representation of the innovations motivates the definition of the pseudo-innovations and hence, the pseudo-Gaussian likelihood function. The term pseudo reflects in the first case that we do not use the real innovations and in the second case that we do not have a Gaussian model. This approach is standard for stationary models (see Schlemm and Stelzer [51]) but it is not so well investigated for non-stationary models. In our model, the pseudo-innovations are as well non-stationary and hence, classical methods for QML estimation for stationary models do not work, e.g., the convergence of the quasi-maximum-likelihood function by a law of large numbers or an ergodic theorem.

Well-known achievements on ML estimation for integrated and cointegrated processes in discrete time are Saikkonen [46, 47]. Under the constraint that the ML estimator is consistent and the long-run parameter estimator satisfies some appropriate order of consistency condition, the papers present stochastic equicontinuity criteria for the standardized score vector and the standardized Hessian matrix such that the asymptotic distribution of the ML estimator can be calculated. The main contributions of these papers are the derivation of stochastic equicontinuity and weak convergence results of various first and second order sample moments from integrated processes. The concepts are applied to a ML estimator in a simple regression model with integrated and stationary regressors.

In this paper, we follow the ideas of Saikkonen [47] to derive the asymptotic distribution of the QML estimator by providing evidence that these three criteria are satisfied. However, our model does not satisfy the stochastic equicontinuity conditions of Saikkonen [46, 47] such that the weak convergence results of these papers cannot be applied directly. But we use a similar approach. In the derivation of the consistency of the QML estimator we even require local Lipschitz continuity for some parts of the likelihood-function which is stronger than local stochastic equicontinuity. For this reason we pay our attention in this paper to local Lipschitz continuity instead of stochastic equicontinuity.

Although Saikkonen [46, 47] presents no general conditions for the analysis of the consistency and the order of consistency of a ML estimator in an integrated or cointegrated model, the verification of the consistency of the ML estimator in the regression example of Saikkonen [47] suggests, how to proceed in more general models. That is done by a stepwise approach: In the first step, we prove the consistency of the long-run parameter estimator and in the second step its consistency rate; the long-run parameter estimator is super-consistent. In the third step, we are able to prove the consistency of the short-run parameter estimator. However, important for the proofs is, as in Saikkonen [47], the appropriate division of the likelihood-function where one part of the likelihood-function depends only on the short-run parameters and is based on stationary processes. This decomposition is not obvious and presumes as well a splitting of the pseudo-innovations in a non-stationary and a stationary part depending only on the short-run parameters.

The paper is structured on the following way. An introduction into QML estimation for cointegrated continuous-time linear state space models is given in Section 2. First, we state in Section 2.1 the assumptions about our parametric family of cointegrated output processes Y. Then, we define the pseudo-innovations for the QML estimation by the Kalman filter in Section 2.2. Based on the pseudo-innovations we calculate the pseudo-Gaussian log-likelihood function in Section 2.3. In Section 2.4 we introduce some identifiability conditions to get a unique minimum of the likelihood function. The main results of this paper are given in Section 3 and Section 4. First, we show the consistency of the QML estimator in Section 3. Next, we calculate the asymptotic distribution of the QML estimator in Section 4. The short-run QML estimator is asymptotically normally distributed and mimics the properties of QML estimators for stationary models. In contrast, the long-run QML estimator is asymptotically mixed normally distributed with a convergence rate of n instead of $\sqrt{n}$ as occurring in stationary models. Finally, in Section 5 we show the performance of our estimator in a simulation study, and in Section 6 we give some conclusions. Eventually, in Appendix A we present some asymptotic results and local Lipschitz continuity conditions which we use throughout the paper. Because of their technicality and to keep the paper readable, they are moved to the appendix.

Notation We use as norms the Euclidean norm $\|\cdot\|$ in $\mathbb{R}^d$ and the Frobenius norm $\|\cdot\|$ for matrices, which is submultiplicative. $0_{d\times s}$ denotes the zero matrix in $\mathbb{R}^{d\times s}$ and I_d is the identity matrix in $\mathbb{R}^{d\times d}$. For a matrix $A \in \mathbb{R}^{d\times d}$ we denote by A^T its transpose, $\operatorname{tr}(A)$ its trace, $\det(A)$ its determinant, rank A its rank, $\lambda_{\min}(A)$ its smallest eigenvalue and $\sigma_{\min}(A)$ its smallest singular value. If A is symmetric and positive semi-definite, we write $A^{\frac{1}{2}}$ for the principal square root, i.e., $A^{\frac{1}{2}}$ is a symmetric, positive semi-definite matrix satisfying

$A^{\frac{1}{2}}A^{\frac{1}{2}} = A$. For a matrix $A \in \mathbb{R}^{d\times s}$ with rank $A = s$, $A^{\perp}$ is a $d \times (d-s)$-dimensional matrix with rank $(d-s)$ satisfying $A^{\mathsf{T}}A^{\perp} = 0_{s\times(d-s)}$ and $A^{\perp\mathsf{T}}A = 0_{(d-s)\times s}$. For two matrices $A \in \mathbb{R}^{d\times s}$ and $B \in \mathbb{R}^{r\times n}$, we denote by $A \otimes B$ the Kronecker product which is an element of $\mathbb{R}^{dr\times sn}$, by $\mathrm{vec}(A)$ the operator which converts the matrix A into a column vector and by $\mathrm{vech}(A)$ the operator which converts a symmetric matrix A into a column vector by vectorizing only the lower triangular part of A. We write ∂_i for the partial derivative operator with respect to the i^{th} coordinate and $\partial_{i,j}$ for the second partial derivative operator with respect to the i^{th} and j^{th} coordinate. Further, for a matrix function $f(\vartheta)$ in $\mathbb{R}^{d\times m}$ with $\vartheta \in \mathbb{R}^s$ the gradient with respect to the parameter vector ϑ is denoted by $\nabla_\vartheta f(\vartheta) = \frac{\partial \mathrm{vec}(f(\vartheta))}{\partial \vartheta^T} \in \mathbb{R}^{dm\times s}$. Let $\xi = (\xi_k)_{k\in\mathbb{N}}$ and $\eta = (\eta_k)_{k\in\mathbb{N}}$ be d-dimensional stochastic processes then $\Gamma_{\xi,\eta}(l) = \mathrm{Cov}(\xi_1, \eta_{1+l})$ and $\Gamma_\xi(l) = \mathrm{Cov}(\xi_1, \xi_{1+l})$, $l \in \mathbb{N}_0$, are the covariance functions. Finally, we denote with $\xrightarrow{w}$ weak convergence and with $\xrightarrow{p}$ convergence in probability. In general $\mathfrak{C}$ denotes a constant which may change from line to line.

2. Step-wise quasi-maximum likelihood estimation

2.1. Parametric model

Let $\Theta \subset \mathbb{R}^s$, $s \in \mathbb{N}$, be a parameter space. We assume that we have a parametric family $(Y_\vartheta)_{\vartheta\in\Theta}$ of solutions of continuous-time cointegrated linear state space models of the form

$$Y_\vartheta(t) = C_{1,\vartheta}Z + C_{1,\vartheta}B_{1,\vartheta}L_\vartheta(t) + Y_{st,\vartheta}(t), \quad t \geq 0, \tag{2.1}$$

where Z is a random starting vector, $L_\vartheta = (L_\vartheta(t))_{t\geq 0}$ is a Lévy process and $Y_{st,\vartheta} = (Y_{st,\vartheta}(t))_{t\geq 0}$ is a stationary solution of the state-space model

$$\begin{aligned} \mathrm{d}X_{st,\vartheta}(t) &= A_{2,\vartheta}X_{st,\vartheta}(t)\mathrm{d}t + B_{2,\vartheta}\mathrm{d}L_\vartheta(t), \\ Y_{st,\vartheta}(t) &= C_{2,\vartheta}X_{st,\vartheta}(t), \end{aligned} \tag{2.2}$$

with $A_{2,\vartheta} \in \mathbb{R}^{(N-c)\times(N-c)}$, $B_{1,\vartheta} \in \mathbb{R}^{c\times m}$, $B_{2,\vartheta} \in \mathbb{R}^{(N-c)\times m}$, $C_{1,\vartheta} \in \mathbb{R}^{d\times c}$ and $C_{2,\vartheta} \in \mathbb{R}^{d\times(N-c)}$ where $c \leq \min(d,m) \leq N$. In the parameterization of the Lévy process L_ϑ only the covariance matrix Σ_ϑ^L of L_ϑ is parameterized.

The parameter vector of the underlying process Y is denoted by ϑ^0, i.e., $(A_2, B_1, B_2, C_1, C_2, L) = (A_{2,\vartheta^0}, B_{1,\vartheta^0}, B_{2,\vartheta^0}, C_{1,\vartheta^0}, C_{2,\vartheta^0}, L_{\vartheta^0})$ where Y_{st} is a stationary solution of the state space model (A_2, B_2, C_2, L). Throughout the paper, we shortly write $(A_{2,\vartheta}, B_{1,\vartheta}, B_{2,\vartheta}, C_{1,\vartheta}, C_{2,\vartheta}, L_\vartheta)$ for the cointegrated state space model with solution Y_ϑ as defined in (2.1). To be more precise we have the following assumptions on our model.

Assumption A. *For any $\vartheta \in \Theta$ the cointegrated state space model $(A_{2,\vartheta}, B_{1,\vartheta}, B_{2,\vartheta}, C_{1,\vartheta}, C_{2,\vartheta}, L_\vartheta)$ satisfies the following conditions:*

(A1) The parameter space Θ is a compact subset of $\mathbb{R}^s$.

(A2) *The true parameter vector ϑ^0 lies in the interior of Θ.*

(A3) *The Lévy process L_ϑ has mean zero and non-singular covariance matrix $\Sigma_\vartheta^L = \mathbb{E}[L_\vartheta(1)L_\vartheta(1)^\mathsf{T}]$. Moreover, there exists a $\delta > 0$ such that $\mathbb{E}\|L_\vartheta(1)\|^{4+\delta} < \infty$ for any $\vartheta \in \Theta$.*

(A4) *The eigenvalues of $A_{2,\vartheta}$ have strictly negative real parts.*

(A5) *The triplet $(A_{2,\vartheta}, B_{2,\vartheta}, C_{2,\vartheta})$ is minimal with McMillan degree $N-c$ (see Hannan and Deistler [24, Chapter 4.2] for the definition of McMillan degree).*

(A6) *The matrices $B_{1,\vartheta} \in \mathbb{R}^{c\times m}$ and $C_{1,\vartheta}\mathbb{R}^{d\times c}$ have full rank $c \leq \min(d, m)$.*

(A7) *The c-dimensional starting random vector Z does not depend on ϑ, $\mathbb{E}\|Z\|^2 < \infty$ and Z is independent of L_ϑ.*

(A8) *The functions $\vartheta \mapsto A_{2,\vartheta}$, $\vartheta \mapsto B_{i,\vartheta}$, $\vartheta \mapsto C_{i,\vartheta}$ for $i \in \{1, 2\}$, $\vartheta \mapsto \Sigma_\vartheta^L$ and $\vartheta_1 \mapsto C_{1,\vartheta_1}^\perp$ are three times continuously differentiable, where $C_{1,\vartheta}^\perp$ is the unique lower triangular matrix with $C_{1,\vartheta}^{\perp\mathsf{T}} C_{1,\vartheta}^\perp = I_{d-c}$ and $C_{1,\vartheta}^{\perp\mathsf{T}} C_{1,\vartheta} = 0_{(d-c)\times c}$.*

(A9) *$A_\vartheta := diag(0_{c\times c}, A_{2,\vartheta}) \in \mathbb{R}^{N\times N}$, $B_\vartheta := (B_{1,\vartheta}^\mathsf{T}, B_{2,\vartheta}^\mathsf{T})^\mathsf{T} \in \mathbb{R}^{N\times m}$, $C_\vartheta := (C_{1,\vartheta}, C_{2,\vartheta}) \in \mathbb{R}^{d\times N}$. Moreover, C_ϑ has full rank $d \leq N$.*

(A10) *For any $\lambda, \lambda' \in \sigma(A_\vartheta) = \sigma(A_{2,\vartheta}) \cup \{0\}$ and any $k \in \mathbb{Z}\backslash\{0\}$: $\lambda - \lambda' \neq 2\pi k/h$ (Kalman-Bertram criterion).*

Remark 2.1.

(i) *(A1) and (A2) are standard assumptions for QML estimation.*

(ii) *Assumption (A3)-(A4) are sufficient assumptions to guarantee that there exists a stationary solution $Y_{st,\vartheta}$ of the state space model (2.2) (see Marquardt and Stelzer [38]).*

(iii) *Due to assumption (A5) the state space representation of $Y_{st,\vartheta}$ in (2.2) with $A_{2,\vartheta} \in \mathbb{R}^{(N-c)\times(N-c)}$, $B_{2,\vartheta} \in \mathbb{R}^{(d-c)\times m}$ and $C_{2,\vartheta} \in \mathbb{R}^{d\times(N-c)}$ is unique up to a change of basis.*

(iv) *We require that c respectively the cointegration rank $r = d - c$ is known in advance to be able to estimate the model adequately. In reality, it is necessary to estimate first the cointegration rank r and obtain from this $c = d - r$. Possibilities to do this is via information criteria.*

(v) *Using the notation in (A9) it is possible to show that Y_ϑ is the solution of the state space model $(A_\vartheta, B_\vartheta, C_\vartheta, L_\vartheta)$. Furthermore, on account of (A5) and (A6), the state space model $(A_\vartheta, B_\vartheta, C_\vartheta)$ is minimal with McMillan degree N (see Fasen-Hartmann and Scholz [18, Lemma 2.4]) and hence, as well unique up to a change of basis. That in combination with (A10) is sufficient that $Y_\vartheta^{(h)} := (Y_\vartheta^{(h)}(k))_{k\in\mathbb{N}_0} := (Y_\vartheta(kh))_{k\in\mathbb{N}_0}$ is a solution of a discrete-time state space model with McMillan degree N as well.*

Furthermore, we assume that the parameter space Θ is a product space of the form $\Theta = \Theta_1 \times \Theta_2$ with $\Theta_1 \subset \mathbb{R}^{s_1}$ and $\Theta_2 \subset \mathbb{R}^{s_2}$, $s = s_1 + s_2$. The vector $\vartheta = (\vartheta_1^\mathsf{T}, \vartheta_2^\mathsf{T})^\mathsf{T} \in \Theta$ is a s-dimensional parameter vector where $\vartheta_1 \in \Theta_1$ and $\vartheta_2 \in \Theta_2$. The idea is that ϑ_1 is the s_1-dimensional vector of long-run parameters modelling the cointegration space and hence, responsible for the cointegration of Y_ϑ. Whereas ϑ_2 is the s_2-dimensional vector of short-run parameters which has no influence on the cointegration of the model. Since the matrix $C_{1,\vartheta}$ is responsible for the cointegration property (see Fasen-Hartmann and Scholz [18, Theorem 3.3]) we parameterize $C_{1,\vartheta}$ with the sub-vector ϑ_1 and use for all the other matrices ϑ_2. In summary, we parameterize the matrices with the following sub-vectors $(A_{2,\vartheta_2}, B_{1,\vartheta_2}, B_{2,\vartheta_2}, C_{1,\vartheta_1}, C_{2,\vartheta_2}, L_{\vartheta_2})$ for $(\vartheta_1, \vartheta_2) \in \Theta_1 \times \Theta_2 = \Theta$.

2.2. Linear and pseudo-innovations

In this section, we define the pseudo-innovations which are essential to define the QML function. Sampling at distance $h > 0$ maps the class of continuous-time state space models to discrete-time state space models. That class of state space models is not in innovation form and hence, we use a result from Fasen-Hartmann and Scholz [18] to calculate the linear innovations $\varepsilon^*_\vartheta(k) = Y_\vartheta(kh) - P_{k-1}Y_\vartheta(kh)$ where P_k is the orthogonal projection onto $\overline{\text{span}}\{Y_\vartheta(lh) : -\infty < l \leq k\}$ where the closure is taken in the Hilbert space of square-integrable random variables with inner product $(Z_1, Z_2) \mapsto \mathbb{E}(Z_1^\mathsf{T} Z_2)$. Thus, $\varepsilon^*_\vartheta(k)$ is orthogonal to the Hilbert space generated by $\overline{\text{span}}\{Y_\vartheta(lh), -\infty < l < k\}$. In our setting, the linear innovations are as follows.

Proposition 2.2 (Fasen-Hartmann and Scholz [18])**.** *Let $\Omega^{(h)}_\vartheta$ be the unique solution of the discrete-time algebraic Riccati equation*

$$\Omega^{(h)}_\vartheta = \mathrm{e}^{A_\vartheta h}\Omega^{(h)}_\vartheta \mathrm{e}^{A^\mathsf{T}_\vartheta h} - \mathrm{e}^{A_\vartheta h}\Omega^{(h)}_\vartheta C^\mathsf{T}_\vartheta \left(C_\vartheta \Omega^{(h)}_\vartheta C^\mathsf{T}_\vartheta\right)^{-1} C_\vartheta \Omega^{(h)}_\vartheta \mathrm{e}^{A^\mathsf{T}_\vartheta h} + \Sigma^{(h)}_\vartheta,$$

where

$$\Sigma^{(h)}_\vartheta = \int_0^h \begin{pmatrix} B_{1,\vartheta}\Sigma^L_\vartheta B^T_{1,\vartheta} & \mathrm{e}^{A_{2,\vartheta}u} B_{2,\vartheta}\Sigma^L_\vartheta B^T_{1,\vartheta} \\ B_{1,\vartheta}\Sigma^L_\vartheta B^T_{2,\vartheta}\mathrm{e}^{A^T_{2,\vartheta}u} & \mathrm{e}^{A_{2,\vartheta}u} B_{2,\vartheta}\Sigma^L_\vartheta B^T_{2,\vartheta}\mathrm{e}^{A^T_{2,\vartheta}u} \end{pmatrix} du,$$

*and $K^{(h)}_\vartheta = \mathrm{e}^{A_\vartheta h}\Omega^{(h)}_\vartheta C^\mathsf{T}_\vartheta \left(C_\vartheta \Omega^{(h)}_\vartheta C^\mathsf{T}_\vartheta\right)^{-1}$ be the steady-state Kalman gain matrix. Then, the **linear innovations** $\varepsilon^*_\vartheta = (\varepsilon^*_\vartheta(k))_{k\in\mathbb{N}}$ of $Y^{(h)}_\vartheta := (Y^{(h)}_\vartheta(k))_{k\in\mathbb{N}} := (Y_\vartheta(kh))_{k\in\mathbb{N}}$ are the unique stationary solution of the state space equation*

$$\begin{aligned} \varepsilon^*_\vartheta(k) &= Y^{(h)}_\vartheta(k) - C_\vartheta X^*_\vartheta(k), \qquad \textit{where} \\ X^*_\vartheta(k) &= (\mathrm{e}^{A_\vartheta h} - K^{(h)}_\vartheta C_\vartheta)X^*_\vartheta(k-1) + K^{(h)}_\vartheta Y^{(h)}_\vartheta(k-1). \end{aligned} \tag{2.3}$$

Moreover, $V^{(h)}_\vartheta = \mathbb{E}(\varepsilon^*_\vartheta(1)\varepsilon^*_\vartheta(1)^\mathsf{T}) = C_\vartheta \Omega^{(h)}_\vartheta C^\mathsf{T}_\vartheta$ is the prediction covariance matrix of the Kalman filter.

We obtain recursively from (2.3)

$$\begin{aligned} \varepsilon^*_\vartheta(k) &= Y^{(h)}_\vartheta(k) - C_\vartheta(\mathrm{e}^{A_\vartheta h} - K^{(h)}_\vartheta C_\vartheta)^{k-1} X^*_\vartheta(1) \\ &\quad - \sum_{j=1}^{k-1} C_\vartheta(\mathrm{e}^{A_\vartheta h} - K^{(h)}_\vartheta C_\vartheta)^{j-1} K^{(h)}_\vartheta Y^{(h)}_\vartheta(k-j). \end{aligned}$$

However, the question arises which choice of $X^*_\vartheta(1)$ of the Kalman recursion results in the stationary $(\varepsilon^*_\vartheta(k))_{k\in\mathbb{N}}$. This we want to elaborate in the following.

Since all eigenvalues of $(\mathrm{e}^{A_\vartheta h} - K_\vartheta^{(h)} C_\vartheta)$ lie inside the unit circle (see Scholz [52, Lemma 4.6.7]) the matrix function

$$\mathsf{l}(z,\vartheta) := I_d - C_\vartheta \sum_{j=1}^{\infty} \left(\mathrm{e}^{A_\vartheta h} - K_\vartheta^{(h)} C_\vartheta\right)^{j-1} K_\vartheta^{(h)} z^j \quad \text{ for } z \in \mathbb{C}$$

is well-defined and due to Fasen-Hartmann and Scholz [18, Lemma 5.7] has the representation as

$$\mathsf{l}(z,\vartheta) = -\alpha(\vartheta) C_{1,\vartheta}^{\perp\,\mathsf{T}} z + \mathsf{k}(z,\vartheta)(1-z)$$

for the linear filter

$$\mathsf{k}(z,\vartheta) := I_d - \sum_{j=1}^{\infty} \mathsf{k}_j(\vartheta) z^j$$

with $\mathsf{k}_j(\vartheta) := \sum_{i=j}^{\infty} C_\vartheta (\mathrm{e}^{A_\vartheta h} - K_\vartheta^{(h)} C_\vartheta)^i K_\vartheta^{(h)} \in \mathbb{R}^{d\times d}$ and a matrix $\alpha(\vartheta) \in \mathbb{R}^{d\times(d-c)}$ with full rank $d-c$. This representation of $\mathsf{l}(z,\vartheta)$ helps us to choose the initial condition $X_\vartheta^*(1)$ in the Kalman recursion appropriate so that the linear innovations $(\varepsilon_\vartheta^*(k))_{k\in\mathbb{N}}$ are really stationary. Therefore, it is important to know that the stationary process $Y_{st,\vartheta}$ can be defined on $\mathbb{R}$ as $Y_{st,\vartheta}(t) = \int_{-\infty}^{t} f_{st,\vartheta}(t-s)\, dL_\vartheta(s)$, $t \in \mathbb{R}$, with $f_{st,\vartheta}(u) = C_{2,\vartheta} \mathrm{e}^{A_{2,\vartheta} u} B_{2,\vartheta} \mathbb{1}_{[0,\infty)}(u)$ and the Levy process $(L_\vartheta(t))_{t\in\mathbb{R}}$ is defined on the negative real-line as $L_\vartheta(t) = \widetilde{L}_\vartheta(-t-)$ for $t<0$ with an independent copy $(\widetilde{L}_\vartheta(t))_{t\geq 0}$ of $(L_\vartheta(t))_{t\geq 0}$. Then, we have an adequate definition of $\Delta Y_\vartheta^{(h)}(k) := Y_\vartheta^{(h)}(k) - Y_\vartheta^{(h)}(k-1)$ for negative values as well as $\Delta Y_\vartheta^{(h)}(k) = \int_{-\infty}^{kh} f_{\Delta,\vartheta}(kh-s)\, dL_\vartheta(s)$, $k \in \mathbb{Z}$, with $f_{\Delta,\vartheta}(u) = f_{st,\vartheta}(u) - f_{st,\vartheta}(u-h) + C_{1,\vartheta} B_{1,\vartheta} \mathbb{1}_{[0,h)}(u)$. As notation, we use B for the backshift operator satisfying $\mathsf{B} Y_\vartheta^{(h)}(k) = Y_\vartheta^{(h)}(k-1)$.

Lemma 2.3. *Let Assumption A hold. Then,*

$$\varepsilon_\vartheta^*(k) = -\Pi(\vartheta) Y_\vartheta^{(h)}(k-1) + k(\mathsf{B},\vartheta)\Delta Y_\vartheta^{(h)}(k), \quad k \in \mathbb{N},$$

where $\Pi(\vartheta) = \alpha(\vartheta) C_{1,\vartheta}^{\perp\,\mathsf{T}}$ *and* $\mathsf{k}(\mathsf{B},\vartheta)\Delta Y_\vartheta^{(h)}(k) = \Delta Y_\vartheta^{(h)}(k) - \sum_{j=1}^{\infty} \mathsf{k}_j(\vartheta) \Delta Y_\vartheta^{(h)}(k-j)$. *The matrix sequence* $(\mathsf{k}_j(\vartheta))_{j\in\mathbb{N}}$ *is uniformly exponentially bounded, i.e., there exist constants* $\mathfrak{C} > 0$ *and* $0 < \rho < 1$ *such that* $\sup_{\vartheta\in\Theta} \|\mathsf{k}_j(\vartheta)\| \leq \mathfrak{C}\rho^j$, $j \in \mathbb{N}$.

Proof. It remains to show that $(\mathsf{k}_j(\vartheta))_{j\in\mathbb{N}}$ is uniformly exponentially bounded. The proof follows in the same line as Schlemm and Stelzer [51, Lemma 2.6] using that all eigenvalues of $(\mathrm{e}^{A_\vartheta h} - K_\vartheta^{(h)} C_\vartheta)$ lie inside the unit circle (see Scholz [52, Lemma 4.6.7]). □

Due to $\Pi(\vartheta) Y_\vartheta^{(h)}(k-1) = \Pi(\vartheta) Y_{\vartheta,st}^{(h)}(k-1)$ we receive from Lemma 2.3

$$\varepsilon_\vartheta^*(k) = -\Pi(\vartheta) Y_{\vartheta,st}^{(h)}(k-1) + \mathsf{k}(\mathsf{B},\vartheta)\Delta Y_\vartheta^{(h)}(k).$$

From this representation we see nicely that $(\varepsilon^*_\vartheta(k))_{k\in\mathbb{N}}$ is indeed a stationary process. Defining $Y^{(h)}_\vartheta$ on the negative integers as

$$\begin{aligned} Y^{(h)}_\vartheta(-k) &= C_{1,\vartheta}Z + Y_{st,\vartheta}(0) - \sum_{j=0}^{k-1}\Delta Y^{(h)}_\vartheta(-j) \\ &= C_{1,\vartheta}Z + L_\vartheta(-kh) + Y^{(h)}_{st,\vartheta}(-k), \quad k\in\mathbb{N}_0, \end{aligned}$$

the initial condition in the Kalman recursion is

$$X^*_\vartheta(1) := \sum_{j=0}^{\infty}(\mathrm{e}^{A_\vartheta h} - K^{(h)}_\vartheta C_\vartheta)^j K^{(h)}_\vartheta Y^{(h)}_\vartheta(-j)$$

so that

$$\varepsilon^*_\vartheta(k) \quad = Y^{(h)}_\vartheta(k) - \textstyle\sum_{j=1}^{\infty} C_\vartheta(\mathrm{e}^{A_\vartheta h} - K^{(h)}_\vartheta C_\vartheta)^{j-1}K^{(h)}_\vartheta Y^{(h)}_\vartheta(k-j).$$

The representation of the linear innovations in Lemma 2.3 motivates the definition of the pseudo-innovations which are going in the likelihood function.

Definition 2.4. *The* ***pseudo-innovations*** *are defined for $k\in\mathbb{N}$ as*

$$\begin{aligned} \varepsilon^{(h)}_k(\vartheta) &= -\Pi(\vartheta)Y^{(h)}_{k-1} + k(\mathsf{B},\vartheta)\Delta Y^{(h)}_k \\ &= Y^{(h)}_k - \sum_{j=1}^{\infty} C_\vartheta(\mathrm{e}^{A_\vartheta h} - K^{(h)}_\vartheta C_\vartheta)^{j-1}K^{(h)}_\vartheta Y^{(h)}_{k-j}. \end{aligned}$$

The main difference of the linear innovations and the pseudo-innovations is that in the linear innovation $Y^{(h)}_\vartheta$ is going in, whereas in the pseudo-innovations $Y^{(h)}$ is going in. For $\vartheta=\vartheta^0$ the pseudo-innovations $(\varepsilon^{(h)}_k(\vartheta^0))_{k\in\mathbb{N}}$ are the linear-innovations $(\varepsilon^*_{\vartheta^0}(k))_{k\in\mathbb{N}}$. In Appendix B we present some probabilistic properties of the pseudo-innovations which we use throughout the paper. In particular, we see that the pseudo-innovations are three times differentiable.

2.3. Quasi-maximum likelihood estimation

We estimate the model parameters via an adapted quasi-maximum likelihood estimation method. Minus two over n times the logarithm of the pseudo-Gaussian likelihood function is given by

$$\mathcal{L}^{(h)}_n(\vartheta) = \frac{1}{n}\sum_{k=1}^{n}\left[d\log 2\pi + \log\det V^{(h)}_\vartheta + \varepsilon^{(h)}_k(\vartheta)^\mathsf{T}\left(V^{(h)}_\vartheta\right)^{-1}\varepsilon^{(h)}_k(\vartheta)\right].$$

The pseudo-innovations $\varepsilon^{(h)}_k(\vartheta)$ are constructed by the infinite past $\{Y^{(h)}(l) : -\infty < l < k\}$. However, the infinite past is not known, we only have the finite observations $Y^{(h)}_1,\ldots,Y^{(h)}_n$. Therefore, we have to approximate the pseudo-innovations and the likelihood-function. For a starting value $\widehat{X}_1^{(h)}(\vartheta)$, which is usually a deterministic constant, we define recursively based on (2.3) the approximate pseudo-innovations as

$$\begin{aligned}\widehat{X}_k^{(h)}(\vartheta) &= (\mathrm{e}^{A_\vartheta h} - K_\vartheta^{(h)} C_\vartheta)\widehat{X}_{k-1}^{(h)}(\vartheta) + K_\vartheta^{(h)} Y_{k-1}^{(h)},\\ \widehat{\varepsilon}_k^{(h)}(\vartheta) &= Y_k^{(h)} - C_\vartheta \widehat{X}_k^{(h)}(\vartheta),\end{aligned}$$

and the approximate likelihood-function as

$$\widehat{\mathcal{L}}_n^{(h)}(\vartheta) = \frac{1}{n}\sum_{k=1}^n \left[d\log 2\pi + \log\det V_\vartheta^{(h)} + \widehat{\varepsilon}_k^{(h)}(\vartheta)^\mathsf{T} \big(V_\vartheta^{(h)}\big)^{-1}\widehat{\varepsilon}_k^{(h)}(\vartheta)\right].$$

Then, the QML estimator

$$\widehat{\vartheta}_n := (\widehat{\vartheta}_{n,1}^\mathsf{T}, \widehat{\vartheta}_{n,2}^\mathsf{T})^\mathsf{T} := \operatorname{argmin}_{\vartheta\in\Theta}\widehat{\mathcal{L}}_n^{(h)}(\vartheta)$$

is defined as the minimizer of the pseudo-Gaussian log-likelihood function $\widehat{\mathcal{L}}_n^{(h)}(\vartheta)$. The estimator $\widehat{\vartheta}_{n,1}$ estimates the long-run parameter ϑ_1 and the estimator $\widehat{\vartheta}_{n,2}$ estimates the short-run parameter ϑ_2. However, for our asymptotic results it does not matter if we use $\widehat{\mathcal{L}}_n^{(h)}(\vartheta)$ or $\mathcal{L}_n^{(h)}(\vartheta)$ as a conclusion of the next proposition. But, for that proposition to hold, we require Assumption B which assumes uniform bounds on the second moments of the starting value $\widehat{X}_1^{(h)}(\vartheta)$ of the Kalman recursion and its partial derivatives.

Assumption B.
For every $u, v \in \{1,\ldots,s\}$ we assume that $\mathbb{E}\big(\sup_{\vartheta\in\Theta}\|\widehat{X}_1^{(h)}(\vartheta)\|^2\big) < \infty$, $\mathbb{E}\left(\sup_{\vartheta\in\Theta}\|\partial_u\widehat{X}_1^{(h)}(\vartheta)\|^2\right) < \infty$, $\mathbb{E}\left(\sup_{\vartheta\in\Theta}\|\partial_{u,v}\widehat{X}_1^{(h)}(\vartheta)\|^2\right) < \infty$ *and* $\widehat{X}_1^{(h)}(\vartheta)$ *is independent of* $(L_\vartheta(t))_{t\geq 0}$.

This assumption is not very restrictive, e.g., if $\widehat{X}_1^{(h)}(\vartheta) = \widehat{X}_1^{(h)}(\vartheta^0)$ for any $\vartheta\in\Theta$ and $\widehat{X}_1^{(h)}(\vartheta^0)$ is a deterministic vector, which we usually have in practice, Assumption B is automatically satisfied.

Proposition 2.5. *Let Assumption A and B hold. Moreover, let $\gamma < 1$ and $u, v \in \{1,\ldots,s\}$. Then,*

(a) $n^\gamma \sup_{\vartheta\in\Theta}|\widehat{\mathcal{L}}_n^{(h)}(\vartheta) - \mathcal{L}_n^{(h)}(\vartheta)| \xrightarrow{p} 0$,

(b) $n^\gamma \sup_{\vartheta\in\Theta}|\partial_u\widehat{\mathcal{L}}_n^{(h)}(\vartheta) - \partial_u\mathcal{L}_n^{(h)}(\vartheta)| \xrightarrow{p} 0$,

(c) $n^\gamma \sup_{\vartheta\in\Theta}|\partial_{u,v}\widehat{\mathcal{L}}_n^{(h)}(\vartheta) - \partial_{u,v}\mathcal{L}_n^{(h)}(\vartheta)| \xrightarrow{p} 0$.

The proof of this proposition is similarly to the proof of Schlemm and Stelzer [50, Lemma 2.7 and Lemma 2.15]. However, they are some essential differences since in their paper $(Y_k^{(h)})_{k\in\mathbb{N}}$ and $(\varepsilon_k^{(h)}(\vartheta))_{k\in\mathbb{N}}$ are stationary sequences where in our setup they are non-stationary. Furthermore, we require different convergence rates. A detailed proof can be found in Appendix C.

We split now the pseudo-innovation sequence based on the decomposition $\vartheta = (\vartheta_1^\mathsf{T}, \vartheta_2^\mathsf{T})^\mathsf{T}$ so that one part is stationary and depends only on ϑ_2:

$$\begin{aligned}
\varepsilon_k^{(h)}(\vartheta) &= \varepsilon_{k,1}^{(h)}(\vartheta) + \varepsilon_{k,2}^{(h)}(\vartheta),\\
\varepsilon_{k,1}^{(h)}(\vartheta) &:= -\left[\Pi(\vartheta_1,\vartheta_2) - \Pi(\vartheta_1^0,\vartheta_2)\right] Y_{k-1}^{(h)} + \left[\mathsf{k}(\mathsf{B},\vartheta_1,\vartheta_2) - \mathsf{k}(\mathsf{B},\vartheta_1^0,\vartheta_2)\right]\Delta Y_k^{(h)},\\
\varepsilon_{k,2}^{(h)}(\vartheta) &:= \varepsilon_{k,2}^{(h)}(\vartheta_2) = -\Pi(\vartheta_1^0,\vartheta_2) Y_{k-1}^{(h)} + \mathsf{k}(\mathsf{B},\vartheta_1^0,\vartheta_2)\Delta Y_k^{(h)}. \qquad (2.4)
\end{aligned}$$

Due to similar calculations as in (B.1) we receive that

$$\Pi(\vartheta_1^0,\vartheta_2) Y_{k-1}^{(h)} = \Pi(\vartheta_1^0,\vartheta_2) Y_{st,k-1}^{(h)}.$$

Hence,

$$\varepsilon_{k,2}^{(h)}(\vartheta_2) = -\Pi(\vartheta_1^0,\vartheta_2) Y_{st,k-1}^{(h)} + \mathsf{k}(\mathsf{B},\vartheta_1^0,\vartheta_2)\Delta Y_k^{(h)}, \quad k \in \mathbb{N}, \qquad (2.5)$$

is indeed stationary. Moreover, $\varepsilon_{k,1}^{(h)}(\vartheta_1^0,\vartheta_2) = 0$ for any $\vartheta_2 \in \Theta_2$ and $k \in \mathbb{N}$. Finally, we separate the log-likelihood function $\mathcal{L}_n^{(h)}(\vartheta)$ in

$$\mathcal{L}_n^{(h)}(\vartheta) = \mathcal{L}_{n,1}^{(h)}(\vartheta) + \mathcal{L}_{n,2}^{(h)}(\vartheta_2),$$

where

$$\begin{aligned}
\mathcal{L}_{n,1}^{(h)}(\vartheta) &:= \mathcal{L}_n^{(h)}(\vartheta_1,\vartheta_2) - \mathcal{L}_n^{(h)}(\vartheta_1^0,\vartheta_2)\\
&= \log\det V_\vartheta^{(h)} - \log\det V_{\vartheta_1^0,\vartheta_2}^{(h)} + \frac{1}{n}\sum_{k=1}^n \varepsilon_{k,1}^{(h)}(\vartheta)^\mathsf{T}\big(V_\vartheta^{(h)}\big)^{-1}\varepsilon_{k,1}^{(h)}(\vartheta)\\
&\quad + \frac{2}{n}\sum_{k=1}^n \varepsilon_{k,1}^{(h)}(\vartheta)^\mathsf{T}\big(V_\vartheta^{(h)}\big)^{-1}\varepsilon_{k,2}^{(h)}(\vartheta_2) + \frac{1}{n}\sum_{k=1}^n \varepsilon_{k,2}^{(h)}(\vartheta_2)^\mathsf{T}\big(V_\vartheta^{(h)}\big)^{-1}\varepsilon_{k,2}^{(h)}(\vartheta_2)\\
&\quad - \frac{1}{n}\sum_{k=1}^n \varepsilon_{k,2}^{(h)}(\vartheta_2)^\mathsf{T}\big(V_{\vartheta_1^0,\vartheta_2}^{(h)}\big)^{-1}\varepsilon_{k,2}^{(h)}(\vartheta_2),\\
\mathcal{L}_{n,2}^{(h)}(\vartheta_2) &:= \mathcal{L}_n^{(h)}(\vartheta_1^0,\vartheta_2)\\
&= d\log 2\pi + \log\det V_{\vartheta_1^0,\vartheta_2}^{(h)} + \frac{1}{n}\sum_{k=1}^n \varepsilon_{k,2}^{(h)}(\vartheta_2)^\mathsf{T}\big(V_{\vartheta_1^0,\vartheta_2}^{(h)}\big)^{-1}\varepsilon_{k,2}^{(h)}(\vartheta_2).
\end{aligned}$$

Obviously, $\mathcal{L}_{n,2}^{(h)}(\vartheta_2)$ depends only on the short-run parameters, whereas $\mathcal{L}_{n,1}^{(h)}(\vartheta)$ depends on all parameters. Furthermore, we have the following relations:

$$\mathcal{L}_{n,1}^{(h)}(\vartheta_1^0,\vartheta_2) = 0 \quad \text{and} \quad \mathcal{L}_n^{(h)}(\vartheta_1^0,\vartheta_2) = \mathcal{L}_{n,2}^{(h)}(\vartheta_2) \quad \text{for any } \vartheta_2 \in \Theta_2. \qquad (2.6)$$

This immediately implies $\mathcal{L}_n^{(h)}(\vartheta^0) = \mathcal{L}_{n,2}^{(h)}(\vartheta_2^0)$. In the remaining part of the paper, we will see that the asymptotic properties of $\widehat{\vartheta}_{n,1}$ are determined by $\mathcal{L}_{n,1}^{(h)}(\vartheta)$, whereas the asymptotic properties of $\widehat{\vartheta}_{n,2}$ are completely determined by $\mathcal{L}_{n,2}^{(h)}(\vartheta_2)$. Since $\mathcal{L}_{n,2}^{(h)}(\vartheta_2)$ is based only on stationary processes it is not surprising that $\widehat{\vartheta}_{n,2}$ exhibits the same asymptotic properties as QML estimators for stationary processes.

2.4. Identifiability

In order to properly estimate our model, we need a unique minimum of the likelihood function and therefore we need some identifiability criteria for the family of stochastic processes $(Y_\vartheta, \vartheta \in \Theta)$. The first assumption guarantees the uniqueness of the long-run parameter ϑ_1^0.

Assumption C. *There exists a constant $\mathfrak{C}^* > 0$ so that*

$$\|C_{1,\vartheta_1}^{\perp\mathsf{T}} C_1\| \geq \mathfrak{C}^* \|\vartheta_1 - \vartheta_1^0\| \quad \textit{for } \vartheta \in \Theta.$$

Remark 2.6.

(i) Without Assumption C we have only that $\|C_{1,\vartheta_1}^{\perp\mathsf{T}} C_1\|$ has a zero in ϑ_1^0 but not that $\|C_{1,\vartheta_1}^{\perp\mathsf{T}} C_1\| \neq 0$ for $\vartheta_1 \neq \vartheta_1^0$. In particular, $\|C_{1,\vartheta_1}^{\perp\mathsf{T}} C_1\| \neq 0$ for $\vartheta_1 \neq \vartheta_1^0$ implies that the space spanned by C_1 and C_{1,ϑ_1} are not the same.

(ii) Due to the Lipschitz-continuity of $C_{1,\vartheta_1}^{\perp\mathsf{T}}$ and $C_{1,\vartheta_1^0}^{\perp\mathsf{T}} C_1 = 0_{(d-c)\times c}$ the upper bound $\|C_{1,\vartheta_1}^{\perp\mathsf{T}} C_1\| \leq \mathfrak{C}\|\vartheta_1 - \vartheta_1^0\|$ for some constant $\mathfrak{C} > 0$ is valid as well.

(iii) Assumption C implies that $\|\Pi(\vartheta)C_1B_1\| = \|\alpha(\vartheta)C_{1,\vartheta_1}^{\perp\mathsf{T}} C_1B_1\| > 0$ for $\vartheta_1^0 \neq \vartheta_1$ since $\alpha(\vartheta)$ and B_1 have full rank, and thus, the process $\left(\varepsilon_{k,1}^{(h)}(\vartheta)\right)_{k\in\mathbb{N}}$ is indeed non-stationary for all long-run parameters $\vartheta_1 \neq \vartheta_1^0$.

(iv) The matrix function $\alpha(\vartheta)$ is continuous and has full column rank $d-c$ so that necessarily $\inf_{\vartheta\in\Theta} \sigma_{\min}(\alpha(\vartheta)) > 0$. Applying Bernstein [9, Corollary 9.6.7] gives for some constant $\mathfrak{C} > 0$:

$$\|\Pi(\vartheta)C_1\| \geq \inf_{\vartheta\in\Theta} \{\sigma_{\min}(\alpha(\vartheta))\} \|C_{1,\vartheta_1}^{\perp\mathsf{T}} C_1\| \geq \mathfrak{C}\|\vartheta_1 - \vartheta_1^0\|.$$

The next assumption guarantees the uniqueness of the short-run parameter ϑ_2^0.

Assumption D. *For any $\vartheta_2^0 \neq \vartheta_2 \in \Theta_2$ there exists a $z \in \mathbb{C}$ such that either*

$$\begin{aligned} C_{\vartheta_1^0,\vartheta_2}\left[I_N - \left(\mathrm{e}^{A_{\vartheta_2}h} - K_{\vartheta_1^0,\vartheta_2}^{(h)} C_{\vartheta_1^0,\vartheta_2}\right)z\right]^{-1} K_{\vartheta_1^0,\vartheta_2}^{(h)} & \\ \neq \quad C\left[I_N - \left(\mathrm{e}^{Ah} - K^{(h)}C\right)z\right]^{-1} K^{(h)} & \\ \textit{or} \quad V_{\vartheta_1^0,\vartheta_2}^{(h)} \quad \neq \quad V^{(h)}. & \end{aligned}$$

Lemma 2.7. *Let Assumption A and D hold. The function $\mathcal{L}_2^{(h)} : \Theta_2 \to \mathbb{R}$ defined by*

$$\mathcal{L}_2^{(h)}(\vartheta_2) := d\log(2\pi) + \log\det V_{\vartheta_1^0,\vartheta_2}^{(h)} + \mathbb{E}\left(\varepsilon_{1,2}^{(h)}(\vartheta_2)^{\mathsf{T}} \left(V_{\vartheta_1^0,\vartheta_2}^{(h)}\right)^{-1} \varepsilon_{1,2}^{(h)}(\vartheta_2)\right) \quad (2.7)$$

has a unique global minimum at ϑ_2^0.

Proof. The proof is analogous to the proof of Lemma 2.10 in Schlemm and Stelzer [51]. □

Without the additional Assumption D we obtain only that $\mathcal{L}_2^{(h)}(\vartheta_2)$ has a minimum in ϑ_2^0 but not that the minimum is unique.

Due to Fasen-Hartmann and Scholz [18, Theorem 3.2] a canonical form for cointegrated state space processes already exists and can be used to construct a model class satisfying Assumption C and Assumption D. Further details are presented in Fasen-Hartmann and Scholz [19]. Moreover, criteria to overcome the aliasing effect (see Blevins [10], Hansen and Sargent [25], McCrorie [40, 41], Phillips [42, 43], Schlemm and Stelzer [51]) are given there.

3. Consistency of the QML estimator

In order to show the consistency of the QML estimator, we follow the ideas of Saikkonen [47] in his regression model. Thus, we prove the consistency in three steps. In the first step, we prove the consistency of the long-run QML estimator $\widehat{\vartheta}_{n,1}$ and next we determine its consistency rate. Thirdly, we prove the consistency of the short-run QML estimator $\widehat{\vartheta}_{n,2}$ by making use of the consistency rate of the long-run QML estimator. Throughout the rest of this paper, we assume that Assumption A–D always hold. Furthermore, we denote by $(W(r))_{0\leq r\leq 1} = ((W_1(r)^\mathsf{T}, W_2(r)^\mathsf{T}, W_3(r)^\mathsf{T})^\mathsf{T})_{0\leq r\leq 1}$ a $(2d+m)$-dimensional Brownian motion with covariance matrix

$$\Sigma_W = \psi(1)\int_0^h \begin{pmatrix} \Sigma_L & \Sigma_L e^{A_2^T u} \\ e^{A_2 u}B_2\Sigma_L & e^{A_2 u}B_2\Sigma_L B_2^T e^{A_2^T u} \end{pmatrix} du\psi(1)^\mathsf{T}, \tag{3.1}$$

where

$$\psi_0 := \begin{pmatrix} C_1B_1 & C_2 \\ 0_{d\times m} & C_2 \\ I_{m\times m} & 0_{m\times N-c} \end{pmatrix}, \psi_j = \begin{pmatrix} 0_{d\times m} & C_2(e^{A_2hj} - e^{A_2h(j-1)}) \\ 0_{d\times m} & C_2 e^{A_2hj} \\ I_{m\times m} & 0_{m\times N-c} \end{pmatrix},$$

$$\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j, \quad z\in\mathbb{C}, \tag{3.2}$$

$(W_i(r))_{0\leq r\leq 1}$, $i=1,2$, are d-dimensional Brownian motions and $(W_3(r))_{0\leq r\leq 1}$ is an m-dimensional Brownian motion.

3.1. Consistency of the long-run QML estimator

To show the consistency for the long-run parameter, Saikkonen [47, p. 903] suggests in his example that it is sufficient to show the following theorem, where $\mathcal{B}(\vartheta_1^0,\delta) := \{\vartheta_1 \in \Theta_1 : \|\vartheta_1 - \vartheta_1^0\| \leq \delta\}$ denotes the closed ball with radius δ around ϑ_1^0, and $\overline{\mathcal{B}}(\vartheta_1^0,\delta) := \Theta_1\backslash\mathcal{B}(\vartheta_1^0,\delta)$ denotes its complement.

Theorem 3.1. *For any $\delta > 0$ we have*

$$\lim_{n\to\infty}\mathbb{P}\left(\inf_{\vartheta\in\overline{\mathcal{B}}(\vartheta_1^0,\delta)\times\Theta_2} \widehat{\mathcal{L}}_n^{(h)}(\vartheta) - \widehat{\mathcal{L}}_n^{(h)}(\vartheta^0) > 0\right) = 1.$$

Corollary 3.2. *In particular, $\widehat{\vartheta}_{n,1} - \vartheta_1^0 = o_p(1)$.*

3.1.1. Proof of Theorem 3.1

The following lemmata are important for the proof of the theorem.

Lemma 3.3. *Let* $L^{(h)} := (L_k^{(h)})_{k\in\mathbb{Z}} := (L(kh))_{k\in\mathbb{Z}}$ *and define*

$$\begin{aligned}
\mathcal{L}_{n,1,1}^{(h)}(\vartheta) &:= \frac{1}{n}\sum_{k=1}^{n}\left[\Pi(\vartheta)C_1B_1L_{k-1}^{(h)}\right]^{\mathsf{T}}\left(V_\vartheta^{(h)}\right)^{-1}\Pi(\vartheta)C_1B_1L_{k-1}^{(h)},\\
\mathcal{L}_{n,1,2}^{(h)}(\vartheta) &:= \mathcal{L}_{n,1}^{(h)}(\vartheta)-\mathcal{L}_{n,1,1}^{(h)}(\vartheta).
\end{aligned}$$

Then, $|\mathcal{L}_{n,1,2}^{(h)}(\vartheta)|\leq \mathfrak{C}\|\vartheta_1-\vartheta_1^0\|U_n$ *for* $\vartheta\in\Theta$ *with* $U_n=1+V_n+Q_n=\mathcal{O}_p(1)$, *and* V_n *and* Q_n *are defined as in Proposition A.3.*

To conclude $\mathcal{L}_{n,1,2}^{(h)}(\cdot,\vartheta_2)$ is local Lipschitz continuous in ϑ_1^0.

Proof of Lemma 3.3. Define

$$\begin{aligned}
\varepsilon_{k,1,1}^{(h)}(\vartheta) &:= -\left[\Pi(\vartheta_1,\vartheta_2)-\Pi(\vartheta_1^0,\vartheta_2)\right]C_1B_1L_{k-1}^{(h)} = -\Pi(\vartheta_1,\vartheta_2)C_1B_1L_{k-1}^{(h)},\\
\varepsilon_{k,1,2}^{(h)}(\vartheta) &:= \varepsilon_{k,1}^{(h)}(\vartheta)-\varepsilon_{k,1,1}^{(h)}(\vartheta)\\
&= -\left[\Pi(\vartheta_1,\vartheta_2)-\Pi(\vartheta_1^0,\vartheta_2)\right]Y_{st,k-1}^{(h)}\\
&\quad+\left[\mathsf{k}(\mathsf{B},\vartheta_1,\vartheta_2)-\mathsf{k}(\mathsf{B},\vartheta_1^0,\vartheta_2)\right]\Delta Y_k^{(h)}.
\end{aligned}$$

Then, $(\varepsilon_{k,1,2}^{(h)}(\vartheta))_{k\in\mathbb{N}}$ is a stationary sequence and $\varepsilon_{k,1}^{(h)}(\vartheta)=\varepsilon_{k,1,1}^{(h)}(\vartheta)+\varepsilon_{k,1,2}^{(h)}(\vartheta)$. First, note that

$$\begin{aligned}
\mathcal{L}_{n,1,2}^{(h)}(\vartheta) &= \log\det V_{\vartheta_1,\vartheta_2}^{(h)}-\log\det V_{\vartheta_1^0,\vartheta_2}^{(h)}\\
&\quad+\frac{2}{n}\sum_{k=1}^{n}\varepsilon_{k,1,1}^{(h)}(\vartheta)^{\mathsf{T}}\left(V_\vartheta^{(h)}\right)^{-1}[\varepsilon_{k,2}^{(h)}(\vartheta_2)+\varepsilon_{k,1,2}^{(h)}(\vartheta)]\\
&\quad+\frac{1}{n}\sum_{k=1}^{n}\varepsilon_{k,1,2}^{(h)}(\vartheta)^{\mathsf{T}}\left(V_\vartheta^{(h)}\right)^{-1}[2\varepsilon_{k,2}^{(h)}(\vartheta_2)+\varepsilon_{k,1,2}^{(h)}(\vartheta)]\\
&\quad+\frac{1}{n}\sum_{k=1}^{n}\varepsilon_{k,2}^{(h)}(\vartheta_2)^{\mathsf{T}}\left(\left(V_\vartheta^{(h)}\right)^{-1}-\left(V_{\vartheta_1^0,\vartheta_2}^{(h)}\right)^{-1}\right)\varepsilon_{k,2}^{(h)}(\vartheta_2).
\end{aligned}$$

In the following, we use Bernstein [9, (2.2.27) and Corollary 9.3.9] to get the upper bound

$$\begin{aligned}
&\left|\frac{1}{n}\sum_{k=1}^{n}\varepsilon_{k,1,1}^{(h)}(\vartheta)^{\mathsf{T}}\left(V_\vartheta^{(h)}\right)^{-1}[\varepsilon_{k,2}^{(h)}(\vartheta)+\varepsilon_{k,1,2}^{(h)}(\vartheta)]\right|\\
&\quad=\left|\frac{1}{n}\sum_{k=1}^{n}\operatorname{tr}\left(\varepsilon_{k,1,1}^{(h)}(\vartheta)^{\mathsf{T}}\left(V_\vartheta^{(h)}\right)^{-1}[\varepsilon_{k,2}^{(h)}(\vartheta)+\varepsilon_{k,1,2}^{(h)}(\vartheta)]\right)\right|
\end{aligned}$$

$$= \left| \operatorname{tr}\left(\left(V_\vartheta^{(h)}\right)^{-1} \frac{1}{n} \sum_{k=1}^{n} [\varepsilon_{k,2}^{(h)}(\vartheta) + \varepsilon_{k,1,2}^{(h)}(\vartheta)] \varepsilon_{k,1,1}^{(h)}(\vartheta)^\mathsf{T} \right) \right|$$
$$\leq \|(V_\vartheta^{(h)})^{-1}\| \left\| \frac{1}{n} \sum_{k=1}^{n} [\varepsilon_{k,2}^{(h)}(\vartheta) + \varepsilon_{k,1,2}^{(h)}(\vartheta)] \varepsilon_{k,1,1}^{(h)}(\vartheta)^\mathsf{T} \right\|.$$

Similarly we find upper bounds for the other terms. Moreover, due to Lemma B.1(b)

$$\begin{aligned}
|\mathcal{L}_{n,1,2}^{(h)}(\vartheta)| \quad \leq \quad & |\log\det V_{\vartheta_1,\vartheta_2}^{(h)} - \log\det V_{\vartheta_1^0,\vartheta_2}^{(h)}| \\
& + \mathfrak{C} \left\| \frac{1}{n} \sum_{k=1}^{n} [\varepsilon_{k,2}^{(h)}(\vartheta) + \varepsilon_{k,1,2}^{(h)}(\vartheta)] \varepsilon_{k,1,1}^{(h)}(\vartheta)^\mathsf{T} \right\| \\
& + \mathfrak{C} \left\| \frac{1}{n} \sum_{k=1}^{n} [2\varepsilon_{k,2}^{(h)}(\vartheta) + \varepsilon_{k,1,2}^{(h)}(\vartheta)] \varepsilon_{k,1,2}^{(h)}(\vartheta)^\mathsf{T} \right\| \\
& + \mathfrak{C} \left\| \left(V_{\vartheta_1,\vartheta_2}^{(h)}\right)^{-1} - \left(V_{\vartheta_1^0,\vartheta_2}^{(h)}\right)^{-1} \right\| \left\| \frac{1}{n} \sum_{k=1}^{n} \varepsilon_{k,2}^{(h)}(\vartheta_2) \varepsilon_{k,2}^{(h)}(\vartheta_2)^\mathsf{T} \right\|.
\end{aligned}$$

Since V_ϑ^{-1} and $\log\det V_\vartheta$ are Lipschitz continuous by Lemma B.1(a), we obtain

$$\begin{aligned}
|\mathcal{L}_{n,1,2}^{(h)}(\vartheta)| \quad \leq \quad & \mathfrak{C} \Bigg(\|\vartheta_1 - \vartheta_1^0\| + \left\| \frac{1}{n} \sum_{k=1}^{n} [\varepsilon_{k,2}^{(h)}(\vartheta) + \varepsilon_{k,1,2}^{(h)}(\vartheta)] \varepsilon_{k,1,1}^{(h)}(\vartheta)^\mathsf{T} \right\| \\
& + \left\| \frac{1}{n} \sum_{k=1}^{n} [2\varepsilon_{k,2}^{(h)}(\vartheta) + \varepsilon_{k,1,2}^{(h)}(\vartheta)] \varepsilon_{k,1,2}^{(h)}(\vartheta)^\mathsf{T} \right\| \\
& + \|\vartheta_1 - \vartheta_1^0\| \left\| \frac{1}{n} \sum_{k=1}^{n} \varepsilon_{k,2}^{(h)}(\vartheta_2) \varepsilon_{k,2}^{(h)}(\vartheta_2)^\mathsf{T} \right\| \Bigg). \qquad (3.3)
\end{aligned}$$

Moreover, $\Pi(\vartheta)$ is Lipschitz continuous as well (see Lemma B.1(a)) and the sequence of matrix functions $(\mathsf{k}_j(\vartheta))_{j\in\mathbb{N}}$ and $(\nabla_\vartheta \mathsf{k}_j(\vartheta))_{j\in\mathbb{N}}$ are exponentially bounded (see Lemma 2.3 and Lemma B.2). Due to (A.4) and $\varepsilon_{k,1,1}^{(h)}(\vartheta_1^0, \vartheta_2) = 0$ we receive

$$\left\| \frac{1}{n} \sum_{k=1}^{n} [\varepsilon_{k,2}^{(h)}(\vartheta) + \varepsilon_{k,1,2}^{(h)}(\vartheta)] \varepsilon_{k,1,1}^{(h)}(\vartheta)^\mathsf{T} \right\| \leq \mathfrak{C} \|\vartheta_1 - \vartheta_1^0\| V_n. \qquad (3.4)$$

Due to (A.6) and $\varepsilon_{k,1,2}^{(h)}(\vartheta_1^0, \vartheta_2) = 0$ we get

$$\left\| \frac{1}{n} \sum_{k=1}^{n} [2\varepsilon_{k,2}^{(h)}(\vartheta) + \varepsilon_{k,1,2}^{(h)}(\vartheta)] \varepsilon_{k,1,2}^{(h)}(\vartheta)^\mathsf{T} \right\| \leq \mathfrak{C} \|\vartheta_1 - \vartheta_1^0\| Q_n. \qquad (3.5)$$

Finally,

$$\left\| \frac{1}{n} \sum_{k=1}^{n} \varepsilon_{k,2}^{(h)}(\vartheta_2) \varepsilon_{k,2}^{(h)}(\vartheta_2)^\mathsf{T} \right\| \leq \mathfrak{C} Q_n \qquad (3.6)$$

as well. Then, (3.3)-(3.6) result in the upper bound

$$|\mathcal{L}_{n,1,2}^{(h)}(\vartheta)| \leq \mathfrak{C}\|\vartheta_1 - \vartheta_1^0\|(1 + V_n + Q_n).$$

A direct consequence of Proposition A.3 is $U_n = 1 + V_n + Q_n = \mathcal{O}_p(1)$. □

Lemma 3.4.

(a) $\sup_{\vartheta_2 \in \Theta_2} |\mathcal{L}_{n,2}^{(h)}(\vartheta_2) - \boldsymbol{\mathcal{L}}_2^{(h)}(\vartheta_2)| \xrightarrow{p} 0$ *as* $n \to \infty$.

(b) $\frac{1}{n}\mathcal{L}_{n,1}^{(h)}(\vartheta) \xrightarrow{w} \int_0^1 \|(V_\vartheta^{(h)})^{-1/2}\Pi(\vartheta)C_1B_1W_3(r)\|^2 \,\mathrm{d}r$ *and the convergence holds in the space of continuous functions on* Θ *with the supremum norm.*

Proof.
(a) is a consequence of Proposition A.1(a) and the continuous mapping theorem.

(b) First, $\sup_{\vartheta\in\Theta} |\frac{1}{n}\mathcal{L}_{n,1,2}(\vartheta)| = o_p(1)$ due to Lemma 3.3 and Θ compact. Second, a conclusion of Proposition A.1(b) and the continuous mapping theorem is that

$$\begin{aligned}
&\frac{1}{n}\mathcal{L}_{n,1,1}^{(h)}(\vartheta)\\
&\quad= \operatorname{tr}\left(\left(V_\vartheta^{(h)}\right)^{-1/2}\Pi(\vartheta)C_1B_1\left(\frac{1}{n^2}\sum_{k=1}^{n} L_{k-1}^{(h)}L_{k-1}^{(h)\mathsf{T}}\right)B_1^\mathsf{T}C_1^\mathsf{T}\Pi(\vartheta)^\mathsf{T}\left(V_\vartheta^{(h)}\right)^{-1/2}\right)\\
&\xrightarrow{w} \operatorname{tr}\left(\left(V_\vartheta^{(h)}\right)^{-1/2}\Pi(\vartheta)C_1B_1\left(\int_0^1 W_3(r)W_3(r)^\mathsf{T}\,dr\right)B_1^\mathsf{T}C_1^\mathsf{T}\Pi(\vartheta)^\mathsf{T}\left(V_\vartheta^{(h)}\right)^{-1/2}\right)\\
&\quad= \int_0^1 \|(V_\vartheta^{(h)})^{-1/2}\Pi(\vartheta)C_1B_1W_3(r)\|^2 \,\mathrm{d}r,
\end{aligned}$$

and the convergence holds in the space of continuous functions on Θ with the supremum norm due to the continuity of $\Pi(\vartheta)$ and $(V_\vartheta^{(h)})^{-1}$ (see Lemma B.1(a)). In the first and in the last equality we used Bernstein [9, 2.2.27] which allows us to permutate matrices in the trace. □

Proof of Theorem 3.1. On the one hand, due to Proposition 2.5

$$\begin{aligned}
&\inf_{\vartheta\in\overline{\mathcal{B}}(\vartheta_1^0,\delta)\times\Theta_2} \widehat{\mathcal{L}}_n^{(h)}(\vartheta) - \widehat{\mathcal{L}}_n^{(h)}(\vartheta^0)\\
&\quad\geq \inf_{\vartheta\in\overline{\mathcal{B}}(\vartheta_1^0,\delta)\times\Theta_2} (\mathcal{L}_n^{(h)}(\vartheta) - \mathcal{L}_n^{(h)}(\vartheta^0)) - 2\sup_{\vartheta\in\Theta} |\widehat{\mathcal{L}}_n^{(h)}(\vartheta) - \mathcal{L}_n^{(h)}(\vartheta)|\\
&\quad= \inf_{\vartheta\in\overline{\mathcal{B}}(\vartheta_1^0,\delta)\times\Theta_2} \left(\mathcal{L}_n^{(h)}(\vartheta) - \mathcal{L}_n^{(h)}(\vartheta^0)\right) + o_p(1).
\end{aligned}$$

On the other hand, due to Lemma 2.7 and Lemma 3.4(a)

$$\left|\inf_{\vartheta_2\in\Theta_2} \mathcal{L}_{n,2}^{(h)}(\vartheta_2) - \mathcal{L}_{n,2}^{(h)}(\vartheta_2^0)\right|$$

$$\begin{aligned}
&\leq \sup_{\vartheta_2\in\Theta_2} |\mathcal{L}_{n,2}^{(h)}(\vartheta_2) - \boldsymbol{\mathcal{L}}_2^{(h)}(\vartheta_2)| + \left|\inf_{\vartheta_2\in\Theta_2} \boldsymbol{\mathcal{L}}_2^{(h)}(\vartheta_2) - \boldsymbol{\mathcal{L}}_2^{(h)}(\vartheta_2^0)\right| \\
&\quad +|\boldsymbol{\mathcal{L}}_2^{(h)}(\vartheta_2^0) - \mathcal{L}_{n,2}^{(h)}(\vartheta_2^0)| \\
&= o_p(1).
\end{aligned}$$

Using (2.6) and the above results we receive

$$\begin{aligned}
&\inf_{\vartheta\in\overline{\mathcal{B}}(\vartheta_1^0,\delta)\times\Theta_2} \widehat{\mathcal{L}}_n^{(h)}(\vartheta) - \widehat{\mathcal{L}}_n^{(h)}(\vartheta^0) \\
&\quad \geq \inf_{\vartheta\in\overline{\mathcal{B}}(\vartheta_1^0,\delta)\times\Theta_2} \left(\mathcal{L}_n^{(h)}(\vartheta) - \mathcal{L}_n^{(h)}(\vartheta^0)\right) + o_p(1) \\
&\quad \geq \inf_{\vartheta\in\overline{\mathcal{B}}(\vartheta_1^0,\delta)\times\Theta_2} \mathcal{L}_{n,1}^{(h)}(\vartheta) + \inf_{\vartheta\in\Theta_2} \left(\mathcal{L}_{n,2}^{(h)}(\vartheta_2) - \mathcal{L}_{n,2}^{(h)}(\vartheta_2^0)\right) + o_p(1) \\
&\quad = \inf_{\vartheta\in\overline{\mathcal{B}}(\vartheta_1^0,\delta)\times\Theta_2} \mathcal{L}_{n,1}^{(h)}(\vartheta) + o_p(1).
\end{aligned}$$

Hence, it suffices to show that for any $\tau > 0$

$$\lim_{n\to\infty} \mathbb{P}\left(\inf_{\vartheta\in\overline{\mathcal{B}}(\vartheta_1^0,\delta)\times\Theta_2} \mathcal{L}_{n,1}^{(h)}(\vartheta) > \tau\right) = 1. \tag{3.7}$$

An application of Lemma 3.4(b) and the continuous mapping theorem yield

$$\inf_{\vartheta\in\overline{\mathcal{B}}(\vartheta_1^0,\delta)\times\Theta_2} \frac{1}{n}\mathcal{L}_{n,1}^{(h)}(\vartheta) \xrightarrow{w} \inf_{\vartheta\in\overline{\mathcal{B}}(\vartheta_1^0,\delta)\times\Theta_2} \int_0^1 \|(V_\vartheta^{(h)})^{-1/2}\Pi(\vartheta)C_1B_1W_3(r)\|^2\,\mathrm{d}r. \tag{3.8}$$

Due to Bernstein [9, Corollary 9.6.7]

$$\begin{aligned}
&\int_0^1 \|(V_\vartheta^{(h)})^{-1/2}\Pi(\vartheta)C_1B_1W_3(r)\|^2\,dr \\
&\qquad \geq \sigma_{\min}((V_\vartheta^{(h)})^{-1}) \int_0^1 \|\Pi(\vartheta)C_1B_1W_3(r)\|^2\,dr.
\end{aligned} \tag{3.9}$$

Moreover,

$$\begin{aligned}
&\int_0^1 \|\Pi(\vartheta)C_1B_1W_3(r)\|^2\,dr \\
&\quad = \int_0^1 \operatorname{tr}\left([B_1W_3(r)]^\mathsf{T}[\Pi(\vartheta)C_1]^\mathsf{T}[\Pi(\vartheta)C_1][B_1W_3(r)]\right)\,dr \\
&\quad = \operatorname{tr}\left([\Pi(\vartheta)C_1]^\mathsf{T}[\Pi(\vartheta)C_1]\int_0^1 [B_1W_3(r)][B_1W_3(r)]^\mathsf{T}\,dr\right),
\end{aligned}$$

where we used Bernstein [9, 2.2.27] to permutate the matrices in the trace. The random matrix $\int_0^1 [B_1W_3(r)]^\mathsf{T}[B_1W_3(r)]\,dr$ is $\mathbb{P}$-a.s. positive definite since

B_1 and the covariance matrix of W_3 have full rank. Hence, there exists an $m \times m$-dimensional symmetric positive random matrix W^* with $W^* W^{*\mathsf{T}} = \int_0^1 [B_1 W_3(r)][B_1 W_3(r)]^{\mathsf{T}}\, dr$. Then, we obtain similarly as above with Bernstein [9, 2.2.27]

$$\int_0^1 \|\Pi(\vartheta) C_1 B_1 W_3(r)\|^2\, dr = \operatorname{tr}\left([W^*]^{\mathsf{T}} [\Pi(\vartheta) C_1]^{\mathsf{T}} [\Pi(\vartheta) C_1] W^*\right) = \|\Pi(\vartheta) C_1 W^*\|^2.$$

Again an application of Bernstein [9, Corollary 9.6.7] and (3.9) yields

$$\begin{aligned}
&\int_0^1 \|(V_\vartheta^{(h)})^{-1/2} \Pi(\vartheta) C_1 B_1 W_3(r)\|^2\, \mathrm{d}r \qquad (3.10)\\
&\quad \geq \sigma_{\min}((V_\vartheta^{(h)})^{-1}) \int_0^1 \|\Pi(\vartheta) C_1 B_1 W_3(r)\|^2\, dr\\
&\quad = \sigma_{\min}((V_\vartheta^{(h)})^{-1}) \|\Pi(\vartheta) C_1 W^*\|^2\\
&\quad \geq \sigma_{\min}((V_\vartheta^{(h)})^{-1}) \sigma_{\min}\left(W^* W^{*\mathsf{T}}\right) \|\Pi(\vartheta) C_1\|^2\\
&\quad = \sigma_{\min}((V_\vartheta^{(h)})^{-1}) \sigma_{\min}\left(\int_0^1 B_1 W_3(r) [B_1 W_3(r)]^{\mathsf{T}}\, dr\right) \|\Pi(\vartheta) C_1\|^2.
\end{aligned}$$

Since $B_1 \int_0^1 W_3(r) W_3(r)^{\mathsf{T}} \mathrm{d}r B_1^{\mathsf{T}}$ is $\mathbb{P}$-a.s. positive definite

$$\sigma_{\min}\left(B_1 \int_0^1 W_3(r) W_3(r)^T \mathrm{d}r B_1^{\mathsf{T}}\right) > 0 \quad \mathbb{P}\text{-a.s.}$$

On the one hand, $\inf_{\vartheta \in \overline{\mathcal{B}}(\vartheta_1^0, \delta) \times \Theta_2} \sigma_{\min}((V_\vartheta^{(h)})^{-1}) > 0$ due Lemma B.1(c). On the other hand, Assumption C (see Remark 2.6) implies that

$$\inf_{\vartheta \in \overline{\mathcal{B}}(\vartheta_1^0, \delta) \times \Theta_2} \|\Pi(\vartheta) C_1\|^2 > \mathfrak{C}^2 \delta^2 > 0.$$

To conclude

$$\inf_{\vartheta \in \overline{\mathcal{B}}(\vartheta_1^0, \delta) \times \Theta_2} \int_0^1 \|(V_\vartheta^{(h)})^{-1/2} \Pi(\vartheta) C_1 B_1 W_3(r)\|^2\, \mathrm{d}r > 0 \quad \mathbb{P}\text{-a.s.},$$

which finally gives with (3.8) that $\inf_{\vartheta \in \overline{\mathcal{B}}(\vartheta_1^0, \delta) \times \Theta_2} \mathcal{L}_{n,1}^{(h)}(\vartheta) \xrightarrow{p} \infty$ and thus, (3.7) is proven. $\square$

3.2. Super-consistency of the long-run QML estimator

From the previous section we already know that the QML estimator $\widehat{\vartheta}_{n,1}$ for the long-run parameter is consistent. In the following, we will calculate its consistency rate. For $0 \leq \gamma < 1$ define the set

$$N_{n,\gamma}(\vartheta_1^0, \delta) := \left\{\vartheta_1 \in \Theta_1 : \|\vartheta_1 - \vartheta_1^0\| \leq \delta n^{-\gamma}\right\}, \quad n \in \mathbb{N}, \qquad (3.11)$$

and $\overline{N}_{n,\gamma}(\vartheta_1^0, \delta) := \Theta_1 \backslash N_{n,\gamma}(\vartheta_1^0, \delta)$ as its complement. As Saikkonen [47, eq. (26)] we receive the consistency rate from the next statement.

Theorem 3.5. *Let $0 \leq \gamma < 1$. For any $\delta > 0$ we have*

$$\lim_{n\to\infty} \mathbb{P}\left(\inf_{\vartheta \in \overline{N}_{n,\gamma}(\vartheta_1^0,\delta)\times\Theta_2} \widehat{\mathcal{L}}_n^{(h)}(\vartheta) - \widehat{\mathcal{L}}_n^{(h)}(\vartheta^0) > 0 \right) = 1.$$

Corollary 3.6. *In particular, $\widehat{\vartheta}_{n,1} - \vartheta_1^0 = o_p(n^{-\gamma})$ for $0 \leq \gamma < 1$.*

3.2.1. Proof of Theorem 3.5

The proof uses the next lemma.

Lemma 3.7. *Let the notation of Lemma 3.3 hold. Then,*

(a) $\mathcal{L}_{n,1,1}^{(h)}(\vartheta) \geq \mathfrak{C}\sigma_{\min}((V_\vartheta^{(h)})^{-1})\|\vartheta_1 - \vartheta_1^0\|^2 \sigma_{\min}\left(\frac{1}{n}\sum_{k=1}^n B_1 L_{k-1}^{(h)}[B_1 L_{k-1}^{(h)}]^\mathsf{T}\right).$

(b) $\mathcal{L}_{n,1,1}^{(h)}(\vartheta) \leq \mathfrak{C}\|(V_\vartheta^{(h)})^{-1}\|\|\vartheta_1 - \vartheta_1^0\|^2 \operatorname{tr}\left(\frac{1}{n}\sum_{k=1}^n B_1 L_{k-1}^{(h)}[B_1 L_{k-1}^{(h)}]^\mathsf{T}\right).$

Proof.
(a) Several applications of Bernstein [9, Corollary 9.6.7] give, similarly as in (3.10),

$$\begin{aligned}
\mathcal{L}_{n,1,1}^{(h)}(\vartheta) &= \frac{1}{n}\sum_{k=1}^n \|(V_\vartheta^{(h)})^{-1/2}\Pi(\vartheta)C_1 B_1 L_{k-1}^{(h)}\|^2 \\
&\geq \sigma_{\min}((V_\vartheta^{(h)})^{-1})\sigma_{\min}\left(\frac{1}{n}\sum_{k=1}^n B_1 L_{k-1}^{(h)}[B_1 L_{k-1}^{(h)}]^\mathsf{T}\right)\|\Pi(\vartheta)C_1\|^2.
\end{aligned}$$

An application of Assumption C (see Remark 2.6) yields (a).

(b) The submultiplicativity of the norm gives

$$\begin{aligned}
\mathcal{L}_{n,1,1}^{(h)}(\vartheta) &= \frac{1}{n}\sum_{k=1}^n \|(V_\vartheta^{(h)})^{-1/2}\Pi(\vartheta)C_1 B_1 L_{k-1}^{(h)}\|^2 \\
&\leq \|(V_\vartheta^{(h)})^{-1/2}\|^2\|\Pi(\vartheta)C_1\|^2 \frac{1}{n}\sum_{k=1}^n \|B_1 L_{k-1}^{(h)}\|^2 \\
&= \|(V_\vartheta^{(h)})^{-1/2}\|^2\|\Pi(\vartheta)C_1\|^2 \operatorname{tr}\left(\frac{1}{n}\sum_{k=1}^n [B_1 L_{k-1}^{(h)}][B_1 L_{k-1}^{(h)}]^\mathsf{T}\right).
\end{aligned}$$

In the last line we applied Bernstein [9, 2.2.27]. Due to $\Pi(\vartheta_1^0,\vartheta_2)C_1 = 0_{d\times c}$ we have

$$\mathcal{L}_{n,1,1}^{(h)}(\vartheta) \leq \|(V_\vartheta^{(h)})^{-1}\|\|\Pi(\vartheta)C_1 - \Pi(\vartheta_1^0,\vartheta_2)C_1\|^2 \operatorname{tr}\left(\frac{1}{n}\sum_{k=1}^n B_1 L_{k-1}^{(h)}[B_1 L_{k-1}^{(h)}]^\mathsf{T}\right).$$

Finally, the Lipschitz continuity of $\Pi(\vartheta)$ and hence, of $\Pi(\vartheta)C_1$ yield the statement. □

A conclusion of Lemma 3.3 and Lemma 3.7 is the local Lipschitz continuity of $\mathcal{L}_{n,1}^{(h)}(\cdot,\vartheta_2)$ in ϑ_1^0. Essential for the proof of Theorem 3.5 is as well the local Lipschitz continuity of $\mathcal{L}_{n,1,2}^{(h)}(\cdot,\vartheta_2)$ in ϑ_1^0.

Proof of Theorem 3.5. Due to Proposition 2.5 the lower bound

$$\begin{aligned}
&\inf_{\vartheta\in\overline{N}_{n,\gamma}(\vartheta_1^0,\delta)\times\Theta_2} n\big(\widehat{\mathcal{L}}_n^{(h)}(\vartheta)-\widehat{\mathcal{L}}_n^{(h)}(\vartheta^0)\big)\\
&\quad\geq \inf_{\vartheta\in\overline{N}_{n,\gamma}(\vartheta_1^0,\delta)\times\Theta_2} n\big(\mathcal{L}_n^{(h)}(\vartheta)-\mathcal{L}_n^{(h)}(\vartheta^0)\big)+o_p(1)\\
&\quad\geq \inf_{\vartheta\in\overline{N}_{n,\gamma}(\vartheta_1^0,\delta)\times\Theta_2} n\mathcal{L}_{n,1}^{(h)}(\vartheta)+\inf_{\vartheta\in\Theta_2} n\big(\mathcal{L}_{n,2}^{(h)}(\vartheta_2)-\mathcal{L}_{n,2}^{(h)}(\vartheta_2^0)\big)+o_p(1)
\end{aligned}$$

holds. We investigate now the second term. Note that $\mathcal{L}_{n,2}^{(h)}(\vartheta_2)$ depends only on the short-run parameters. Therefore, we take the infeasible estimator

$$\widehat{\vartheta}_{n,2}^{st}:=\arg\min_{\vartheta_2\in\Theta_2}\mathcal{L}_{n,2}^{(h)}(\vartheta_2)$$

for the short-run parameter ϑ_2^0 minimizing $\mathcal{L}_{n,2}^{(h)}(\vartheta_2)$. For this reason, we can interpret this as a "classical" stationary estimation problem. Applying a Taylor-expansion of $n\mathcal{L}_{n,2}^{(h)}$ around ϑ_2^0 yields

$$n\cdot\big(\mathcal{L}_{n,2}^{(h)}(\widehat{\vartheta}_{n,2}^{st})-\mathcal{L}_{n,2}^{(h)}(\vartheta_2^0)\big)=\big(\sqrt{n}\nabla_{\vartheta_2}\mathcal{L}_{n,2}^{(h)}(\underline{\vartheta}_{n,2})\big)\cdot\big(\sqrt{n}(\widehat{\vartheta}_{n,2}^{st}-\vartheta_2^0)\big)$$

for an appropriate intermediate value $\underline{\vartheta}_{n,2}\in\Theta_2$ with $\|\underline{\vartheta}_{n,2}-\vartheta_2^0\|\leq\|\widehat{\vartheta}_{n,2}^{st}-\vartheta_2^0\|$. Since $\sqrt{n}\nabla_{\vartheta_2}\mathcal{L}_{n,2}^{(h)}(\underline{\vartheta}_{n,2})$ and $\sqrt{n}(\widehat{\vartheta}_{n,2}^{st}-\vartheta_2^0)$ are asymptotically normally distributed (these are special and easier calculations as in Section 4.2) we can conclude

$$n\cdot\big(\mathcal{L}_{n,2}^{(h)}(\widehat{\vartheta}_{n,2}^{st})-\mathcal{L}_{n,2}^{(h)}(\vartheta_2^0)\big)=\mathcal{O}_p(1).$$

Finally,

$$\inf_{\vartheta\in\overline{N}_{n,\gamma}(\vartheta_1^0,\delta)\times\Theta_2} n\cdot\big(\widehat{\mathcal{L}}_n^{(h)}(\vartheta)-\widehat{\mathcal{L}}_n^{(h)}(\vartheta^0)\big)\geq\inf_{\vartheta\in\overline{N}_{n,\gamma}(\vartheta_1^0,\delta)\times\Theta_2} n\cdot\mathcal{L}_{n,1}^{(h)}(\vartheta)+\mathcal{O}_p(1).$$

Thus, if we can show that

$$\sup_{\vartheta\in\overline{N}_{n,\gamma}(\vartheta_1^0,\delta)\times\Theta_2} n\mathcal{L}_{n,1}^{(h)}(\vartheta)\xrightarrow{p}\infty, \tag{3.12}$$

then for any $\tau>0$

$$\begin{aligned}
&\lim_{n\to\infty}\mathbb{P}\bigg(\inf_{\vartheta\in\overline{N}_{n,\gamma}(\vartheta_1^0,\delta)\times\Theta_2}\widehat{\mathcal{L}}_n^{(h)}(\vartheta)-\widehat{\mathcal{L}}_n^{(h)}(\vartheta^0)>0\bigg)\\
&\qquad\geq\lim_{n\to\infty}\mathbb{P}\bigg(\inf_{\vartheta\in\overline{N}_{n,\gamma}(\vartheta_1^0,\delta)\times\Theta_2} n\left(\widehat{\mathcal{L}}_n^{(h)}(\vartheta)-\widehat{\mathcal{L}}_n^{(h)}(\vartheta^0)\right)>\tau\bigg)=1.
\end{aligned}$$

Before we prove (3.12) we first note that due to (3.7) we only have to consider the set

$$\overline{M}_{n,\gamma}(\vartheta_1^0,\delta_1) := \overline{N}_{n,\gamma}(\vartheta_1^0,\delta_1) \cap \mathcal{B}(\vartheta_1^0,\delta_1) \subseteq \Theta_1 \cap \mathcal{B}(\vartheta_1^0,\delta_1)$$

for n large enough instead of the whole set $\overline{N}_{n,\gamma}(\vartheta_1^0,\delta_1)$ in the infimum. Note that $\inf_{\vartheta\in\Theta}\sigma_{\min}((V_\vartheta^{(h)})^{-1})>0$ by Lemma B.1(c). Then, Lemma 3.3 and Lemma 3.7 give the lower bound

$$\begin{aligned}
&\mathcal{L}_{n,1}^{(h)}(\vartheta)\\
&\quad\geq\ \mathcal{L}_{n,1,1}^{(h)}(\vartheta)-|\mathcal{L}_{n,1,2}^{(h)}(\vartheta)|\\
&\quad\geq\ \mathfrak{C}\|\vartheta_1-\vartheta_1^0\|^2\sigma_{\min}\left(\frac{1}{n}\sum_{k=1}^n B_1L_{k-1}^{(h)}[B_1L_{k-1}^{(h)}]^\mathsf{T}\right)-\mathfrak{C}\|\vartheta_1-\vartheta_1^0\|U_n\\
&\quad\geq\ \mathfrak{C}n\|\vartheta_1-\vartheta_1^0\|^2\underbrace{\left(\sigma_{\min}\left(\frac{1}{n^2}\sum_{k=1}^n B_1L_{k-1}^{(h)}[B_1L_{k-1}^{(h)}]^\mathsf{T}\right)-\frac{1}{n\|\vartheta_1-\vartheta_1^0\|}U_n\right)}_{=:Z_n(\vartheta)}.
\end{aligned}$$

Finally,

$$\begin{aligned}
&\inf_{\vartheta\in\overline{M}_{n,\gamma}(\vartheta_1^0,\delta)\times\Theta_2} n\mathcal{L}_{n,1}^{(h)}(\vartheta)\\
&\qquad\geq\ \left(\inf_{\vartheta\in\overline{M}_{n,\gamma}(\vartheta_1^0,\delta)\times\Theta_2}\mathfrak{C}n^2\|\vartheta_1-\vartheta_1^0\|^2\right)\left(\inf_{\vartheta\in\overline{M}_{n,\gamma}(\vartheta_1^0,\delta)\times\Theta_2}Z_n(\vartheta)\right)\\
&\qquad\geq\ \mathfrak{C}n^{2-2\gamma}\inf_{\vartheta\in\overline{M}_{n,\gamma}(\vartheta_1^0,\delta)\times\Theta_2}Z_n(\vartheta).
\end{aligned}$$

Due to Proposition A.1(b) and Lemma 3.3, we receive

$$\inf_{\vartheta\in\overline{M}_{n,\gamma}(\vartheta_1^0,\delta)\times\Theta_2}Z_n(\vartheta)\xrightarrow{w}\sigma_{\min}\left(B_1\int_0^1 W_3(r)W_3(r)^\mathsf{T}\mathrm{d}rB_1^\mathsf{T}\right)$$

where the right hand side is almost surely positive. Thus, finally $\sup_{\vartheta\in\overline{M}_{n,\gamma}(\vartheta_1^0,\delta)\times\Theta_2} n\mathcal{L}_{n,1}^{(h)}(\vartheta)\xrightarrow{p}\infty$ for $0\leq\gamma<1$. □

3.3. Consistency of the short-run QML estimator

Next, we consider the consistency of the short-run parameter estimator $\widehat{\vartheta}_{n,2}$ with the help of the order of consistency of the long-run parameter estimator $\widehat{\vartheta}_{n,1}$ which we determined in Corollary 3.6. Similarly to Saikkonen [47, eq. (31)] we show a sufficient condition given by the next theorem. Therefore, define for $\delta>0$ the set $\mathcal{B}(\vartheta_2^0,\delta):=\{\vartheta_2\in\Theta_2:\|\vartheta_2-\vartheta_2^0\|\leq\delta\}$ as closed ball with radius δ around ϑ_2^0 and $\overline{\mathcal{B}}(\vartheta_2^0,\delta):=\Theta_2\backslash\mathcal{B}(\vartheta_2^0,\delta)$ as its complement.

Theorem 3.8. *Then, for any $\delta > 0$ we have*

$$\lim_{n\to\infty} \mathbb{P}\Bigg(\inf_{\vartheta\in\Theta_1\times\overline{B}(\vartheta_2^0,\delta)} \widehat{\mathcal{L}}_n^{(h)}(\vartheta) - \widehat{\mathcal{L}}_n^{(h)}(\vartheta^0) > 0 \Bigg) = 1.$$

Corollary 3.9. *In particular, $\widehat{\vartheta}_{n,2} - \vartheta_2^0 = o_p(1)$.*

3.3.1. Proof of Theorem 3.8

Again we prove some auxiliary results before we state the proof of the theorem. Lemma 3.10 corresponds to Saikkonen [47, eq. (32)] and Lemma 3.11 to Saikkonen [47, eq. (33)] for the regression model.

Lemma 3.10. *For $\frac{1}{2} < \gamma < 1$, $\delta_1 > 0$ and $\tau > 0$ we have*

$$\lim_{n\to\infty} \mathbb{P}\Bigg(\sup_{\vartheta\in N_{n,\gamma}(\vartheta_1^0,\delta_1)\times\Theta_2} |\mathcal{L}_{n,1}^{(h)}(\vartheta)| \leq \tau \Bigg) = 1.$$

Proof. Due to Lemma 3.3 and Lemma 3.7 we have the upper bound

$$\begin{aligned} &|\mathcal{L}_{n,1}^{(h)}(\vartheta)| \\ &\quad\leq\ |\mathcal{L}_{n,1,1}^{(h)}(\vartheta)| + |\mathcal{L}_{n,1,2}^{(h)}(\vartheta)| \\ &\quad\leq\ \mathfrak{C}\|(V_\vartheta^{(h)})^{-1}\|\|\vartheta_1-\vartheta_1^0\|^2 \operatorname{tr}\left(\frac{1}{n}\sum_{k=1}^n B_1 L_{k-1}^{(h)}[B_1 L_{k-1}^{(h)}]^\mathsf{T}\right) + \mathfrak{C}\|\vartheta_1-\vartheta_1^0\| U_n. \end{aligned}$$

Then, Lemma B.1(b) results in

$$\sup_{\vartheta\in N_{n,\gamma}(\vartheta_1^0,\delta_1)\times\Theta_2} \mathcal{L}_{n,1}^{(h)}(\vartheta) \leq \mathfrak{C}\delta_1^2 n^{1-2\gamma} \operatorname{tr}\left(\frac{1}{n^2}\sum_{k=1}^n B_1 L_{k-1}^{(h)}[B_1 L_{k-1}^{(h)}]^\mathsf{T}\right) + \mathfrak{C}\delta_1 n^{-\gamma} U_n. \tag{3.13}$$

Since $U_n = \mathcal{O}_p(1)$ by Lemma 3.3, and

$$\operatorname{tr}\left(\frac{1}{n^2}\sum_{k=1}^n B_1 L_{k-1}^{(h)}[B_1 L_{k-1}^{(h)}]^\mathsf{T}\right) \xrightarrow{w} \operatorname{tr}\left(B_1\int_0^1 W_3(r)W_3(r)^\mathsf{T}\mathrm{d}r B_1^\mathsf{T}\right)$$

by Proposition A.1(b) and the continuous mapping theorem, the right hand side of (3.13) converges to 0 in probability if $\frac{1}{2} < \gamma < 1$. This proves the lemma. □

Lemma 3.11. *For any $\delta > 0$ and $\tau > 0$ we have*

$$\lim_{n\to\infty} \mathbb{P}\Bigg(\inf_{\vartheta_2\in\overline{B}(\vartheta_2^0,\delta)} \mathcal{L}_{n,2}^{(h)}(\vartheta_2) - \mathcal{L}_{n,2}^{(h)}(\vartheta_2^0) > \tau \Bigg) = 1.$$

Proof. We have

$$\begin{aligned}
&\inf_{\vartheta_2\in\overline{\mathcal{B}}(\vartheta_2^0,\delta)}\Big(\mathcal{L}_{n,2}^{(h)}(\vartheta_2)-\mathcal{L}_{n,2}^{(h)}(\vartheta_2^0)\Big)\\
&\qquad\geq \inf_{\vartheta_2\in\overline{\mathcal{B}}(\vartheta_2^0,\delta)}\Big(\mathcal{L}_{n,2}^{(h)}(\vartheta_2)-\boldsymbol{\mathcal{L}}_{2}^{(h)}(\vartheta_2)\Big)+\inf_{\vartheta_2\in\overline{\mathcal{B}}(\vartheta_2^0,\delta)}\Big(-\mathcal{L}_{n,2}^{(h)}(\vartheta_2^0)+\boldsymbol{\mathcal{L}}_{2}^{(h)}(\vartheta_2^0)\Big)\\
&\qquad\quad+\inf_{\vartheta_2\in\overline{\mathcal{B}}(\vartheta_2^0,\delta)}\Big(\boldsymbol{\mathcal{L}}_{2}^{(h)}(\vartheta_2)-\boldsymbol{\mathcal{L}}_{2}^{(h)}(\vartheta_2^0)\Big).
\end{aligned}$$

On the one hand, the first two terms converge to zero in probability, due to Lemma 3.4(a) and the continuous mapping theorem. On the other hand,

$$\inf_{\vartheta_2\in\overline{\mathcal{B}}(\vartheta_2^0,\delta)}\big(\boldsymbol{\mathcal{L}}_{2}^{(h)}(\vartheta_2)-\boldsymbol{\mathcal{L}}_{2}^{(h)}(\vartheta_2^0)\big)>0$$

since $\boldsymbol{\mathcal{L}}_{2}^{(h)}(\vartheta_2)$ has a unique minimum in ϑ_2^0 by Lemma 2.7. □

Proof of Theorem 3.8. Let us assume that $\frac{1}{2}<\gamma<1$. Apparently, the parameter subspace Θ_1 is the union of $\Theta_1=\overline{N}_{n,\gamma}(\vartheta_1^0,\underline{\delta_1})\cup N_{n,\gamma}(\vartheta_1^0,\delta_1)$ and thus, we have already shown Theorem 3.8 for the set $\overline{N}_{n,\gamma}(\vartheta_1^0,\delta_1)\times\overline{\mathcal{B}}(\vartheta_2^0,\delta)$ instead of $\Theta_1\times\overline{\mathcal{B}}(\vartheta_2^0,\delta)$ in Theorem 3.5. It remains to investigate the set $N_{n,\gamma}(\vartheta_1^0,\delta_1)\times\overline{\mathcal{B}}(\vartheta_2^0,\delta)$. For any $\delta_1>0$ we obtain by Proposition 2.5

$$\begin{aligned}
&\lim_{n\to\infty}\mathbb{P}\Bigg(\inf_{\vartheta\in N_{n,\gamma}(\vartheta_1^0,\delta_1)\times\overline{\mathcal{B}}(\vartheta_2^0,\delta)}\big(\widehat{\mathcal{L}}_n^{(h)}(\vartheta)-\widehat{\mathcal{L}}_n^{(h)}(\vartheta^0)\big)>0\Bigg)\\
&=\lim_{n\to\infty}\mathbb{P}\Bigg(\inf_{\vartheta\in N_{n,\gamma}(\vartheta_1^0,\delta_1)\times\overline{\mathcal{B}}(\vartheta_2^0,\delta)}\big(\mathcal{L}_n^{(h)}(\vartheta)-\mathcal{L}_n^{(h)}(\vartheta^0)\big)>0\Bigg)\\
&\geq\lim_{n\to\infty}\mathbb{P}\Bigg(\inf_{\vartheta\in N_{n,\gamma}(\vartheta_1^0,\delta_1)\times\overline{\mathcal{B}}(\vartheta_2^0,\delta)}\mathcal{L}_{n,1}^{(h)}(\vartheta)+\inf_{\vartheta_2\in\overline{\mathcal{B}}(\vartheta_2^0,\delta)}\big(\mathcal{L}_{n,2}^{(h)}(\vartheta_2)-\mathcal{L}_{n,2}^{(h)}(\vartheta_2^0)\big)>0\Bigg)\\
&\geq\lim_{n\to\infty}\mathbb{P}\Bigg(\sup_{\vartheta\in N_{n,\gamma}(\vartheta_1^0,\delta_1)\times\overline{\mathcal{B}}(\vartheta_2^0,\delta)}|\mathcal{L}_{n,1}^{(h)}(\vartheta)|\leq\tau;\inf_{\vartheta_2\in\overline{\mathcal{B}}(\vartheta_2^0,\delta)}\mathcal{L}_{n,2}^{(h)}(\vartheta_2)-\mathcal{L}_{n,2}^{(h)}(\vartheta_2^0)>\tau\Bigg).
\end{aligned}$$

Then, a consequence of Lemma 3.10 and Lemma 3.11 is

$$\lim_{n\to\infty}\mathbb{P}\Bigg(\inf_{\vartheta\in N_{n,\gamma}(\vartheta_1^0,\delta_1)\times\overline{\mathcal{B}}(\vartheta_2^0,\delta)}\big(\widehat{\mathcal{L}}_n^{(h)}(\vartheta)-\widehat{\mathcal{L}}_n^{(h)}(\vartheta^0)\big)>0\Bigg)\geq 1,$$

which proves in combination with Theorem 3.5 the claim. □

4. Asymptotic distributions of the QML estimator

The aim of this section is to derive the asymptotic distributions of the long-run parameter estimator $\widehat{\vartheta}_{n,1}$ and the short-run parameter estimator $\widehat{\vartheta}_{n,2}$. These two estimators have a different asymptotic behavior and a different convergence rate. On the one hand, we prove the asymptotic normality of the short-run QML estimator and on the other hand, we show that the long-run QML estimator is asymptotically mixed normally distributed.

4.1. Asymptotic distribution of the long-run parameter estimator

We derive in this section the asymptotic distribution of the long-run QML estimator $\widehat{\vartheta}_{n,1}$. From Corollary 3.6 we already know that $\widehat{\vartheta}_{n,1} - \vartheta_1^0 = o_p(n^{-\gamma})$, for $0 \leq \gamma < 1$. Since the true parameter $\vartheta^0 = ((\vartheta_1^0)^\mathsf{T}, (\vartheta_2^0)^\mathsf{T})^\mathsf{T}$ is an element of the interior of the compact parameter space $\Theta = \Theta_1 \times \Theta_2$ due to Assumption A, the estimator $\widehat{\vartheta}_{n,1}$ is at some point also an element of the interior of Θ_1 with probability one. Because the parametrization is assumed to be threefold continuously differentiable, we can find the minimizing $\widehat{\vartheta}_n = (\widehat{\vartheta}_{n,1}^\mathsf{T}, \widehat{\vartheta}_{n,2}^\mathsf{T})^\mathsf{T}$ via the first order condition $\nabla_{\vartheta_1} \widehat{\mathcal{L}}_n^{(h)}(\widehat{\vartheta}_{n,1}, \widehat{\vartheta}_{n,2}) = 0_{s_1}$. We apply a Taylor-expansion of the score vector around the point $(\vartheta_1^0, \widehat{\vartheta}_{n,2})$ resulting in

$$0_{s_1} = \nabla_{\vartheta_1} \widehat{\mathcal{L}}_n^{(h)}(\vartheta_1^0, \widehat{\vartheta}_{n,2}) + n^{-1} \nabla_{\vartheta_1}^2 \underline{\widehat{\mathcal{L}}}_n^{(h)}(\underline{\vartheta}_{n,1}, \widehat{\vartheta}_{n,2}) n(\widehat{\vartheta}_{n,1} - \vartheta_1^0), \tag{4.1}$$

where $\nabla_{\vartheta_1}^2 \underline{\widehat{\mathcal{L}}}_n^{(h)}(\underline{\vartheta}_{n,1}, \widehat{\vartheta}_{n,2})$ denotes the matrix whose i^{th} row, $i = 1, \ldots, s_1$, is equal to the i^{th} row of $\nabla_{\vartheta_1}^2 \widehat{\mathcal{L}}_n^{(h)}(\underline{\vartheta}_{n,1}^i, \widehat{\vartheta}_{n,2})$ with $\underline{\vartheta}_{n,1}^i \in \Theta_1$ such that $\|\underline{\vartheta}_{n,1}^i - \vartheta_1^0\| \leq \|\widehat{\vartheta}_{n,1} - \vartheta_1^0\|$. In the case $\nabla_{\vartheta_1}^2 \underline{\widehat{\mathcal{L}}}_n^{(h)}(\underline{\vartheta}_{n,1}, \widehat{\vartheta}_{n,2})$ is non-singular we receive

$$n(\widehat{\vartheta}_{n,1} - \vartheta_1^0) = -\Big(n^{-1} \nabla_{\vartheta_1}^2 \underline{\widehat{\mathcal{L}}}_n^{(h)}(\underline{\vartheta}_{n,1}, \widehat{\vartheta}_{n,2})\Big)^{-1} \nabla_{\vartheta_1} \widehat{\mathcal{L}}_n^{(h)}(\vartheta_1^0, \widehat{\vartheta}_{n,2}).$$

Thus, we have to consider the asymptotic behavior of the score vector $\nabla_{\vartheta_1} \widehat{\mathcal{L}}_n^{(h)}(\vartheta)$ and the Hessian matrix $\nabla_{\vartheta_1}^2 \widehat{\mathcal{L}}_n^{(h)}(\vartheta)$. Based on Proposition 2.5 it is sufficient to consider $\nabla_{\vartheta_1}^2 \mathcal{L}_n^{(h)}(\vartheta)$ and $\nabla_{\vartheta_1} \mathcal{L}_n^{(h)}(\vartheta)$, respectively.

4.1.1. Asymptotic behavior of the score vector

First, we show the convergence of the gradient with respect to the long-run parameter ϑ_1. For this, we consider the partial derivatives with respect to the i^{th}-component of the parameter vector ϑ, $i = 1, \ldots, s_1$, of the log-likelihood function. These partial derivatives are given due to differentiation rules for matrix functions (see, e.g., Lütkepohl [35, Appendix A.13]) by

$$\begin{aligned}
\partial_i \mathcal{L}_n^{(h)}(\vartheta) = {} & \operatorname{tr}\Big(\big(V_\vartheta^{(h)}\big)^{-1} \partial_i V_\vartheta^{(h)} \Big) \\
& - \frac{1}{n} \sum_{k=1}^{n} \operatorname{tr}\Big(\big(V_\vartheta^{(h)}\big)^{-1} \varepsilon_k^{(h)}(\vartheta) \varepsilon_k^{(h)}(\vartheta)^\mathsf{T} \big(V_\vartheta^{(h)}\big)^{-1} \partial_i V_\vartheta^{(h)} \Big) \\
& + \frac{2}{n} \sum_{k=1}^{n} \big(\partial_i \varepsilon_k^{(h)}(\vartheta)^\mathsf{T}\big) \big(V_\vartheta^{(h)}\big)^{-1} \varepsilon_k^{(h)}(\vartheta).
\end{aligned} \tag{4.2}$$

From Appendix B we already know that the pseudo-innovations are indeed three times differentiable.

For reasons of brevity, we write $\partial_i^1 := \frac{\partial}{\partial \vartheta_{1i}}$ for the partial derivatives with respect to the i^{th}-component of the long-run parameter vector $\vartheta_1 \in \Theta_1$, $i \in \{1, \ldots, s_1\}$, and similarly $\partial_j^{st} := \frac{\partial}{\partial \vartheta_{2j}}$ for the partial derivatives with respect to the j^{th}-component of the short-run parameter vector $\vartheta_2 \in \Theta_2$, $j \in \{1, \ldots, s_2\}$. Analogously we define $\partial_{i,j}^1$ and $\partial_{i,j}^{st}$, respectively for the second partial derivatives.

Proposition 4.1. *The score vector with respect to the long-run parameter ϑ_1 satisfies*

$$\nabla_{\vartheta_1} \mathcal{L}_n^{(h)}(\vartheta^0) \xrightarrow{w} \mathcal{J}_1(\vartheta^0) := \left(\mathcal{J}_1^{(1)}(\vartheta^0) \quad \cdots \quad \mathcal{J}_1^{(s_1)}(\vartheta^0)\right)^{\mathsf{T}},$$

where

$$\begin{aligned} &\mathcal{J}_1^{(i)}(\vartheta^0) \\ &\quad = \quad 2\,\mathrm{tr}\left[\left(V_{\vartheta^0}^{(h)}\right)^{-1}\left(-\partial_i^1 \Pi(\vartheta^0), 0_{d\times d}\right)\int_0^1 W^{\#}(r)\,\mathrm{d}W^{\#}(r)^{\mathsf{T}}\begin{pmatrix} \mathsf{k}(1,\vartheta^0) \\ -\Pi(\vartheta^0)\end{pmatrix}\right] \\ &\qquad\quad +2\,\mathrm{tr}\left[\left(V_{\vartheta^0}^{(h)}\right)^{-1}\left(\Gamma_{\partial_i^1 \mathsf{k}(\mathsf{B},\vartheta^0)\Delta Y^{(h)},\varepsilon^{(h)}(\vartheta^0)}(0)\right)\right] \\ &\qquad\quad +2\,\mathrm{tr}\left[\left(V_{\vartheta^0}^{(h)}\right)^{-1}\left(\sum_{j=1}^{\infty}\Gamma_{-\partial_i^1 \Pi(\vartheta^0)\Delta Y^{(h)},\varepsilon^{(h)}(\vartheta^0)}(j)\right)\right] \end{aligned}$$

and $(W^{\#}(r))_{0\le r\le 1} = ((W_1(r)^{\mathsf{T}}, W_2(r)^{\mathsf{T}})^{\mathsf{T}})_{0\le r\le 1}$ *is defined on p.* *5166*.

Proof. Equation (4.2) implies for $i = 1, \ldots, s_1$ that

$$\begin{aligned} \partial_i^1 \mathcal{L}_n^{(h)}(\vartheta^0) \quad &= \quad \mathrm{tr}\left(\left(V_{\vartheta^0}^{(h)}\right)^{-1}\partial_i^1 V_{\vartheta^0}^{(h)}\right) \\ &\qquad - \mathrm{tr}\left(\left(V_{\vartheta^0}^{(h)}\right)^{-1}\left(\partial_i^1 V_{\vartheta^0}^{(h)}\right)\left(V_{\vartheta^0}^{(h)}\right)^{-1}\frac{1}{n}\sum_{k=1}^{n}\varepsilon_k^{(h)}(\vartheta^0)\varepsilon_k^{(h)}(\vartheta^0)^{\mathsf{T}}\right) \\ &\qquad +2\,\mathrm{tr}\left(\left(V_{\vartheta^0}^{(h)}\right)^{-1}\frac{1}{n}\sum_{k=1}^{n}\left(\partial_i^1 \varepsilon_k^{(h)}(\vartheta^0)\right)\varepsilon_k^{(h)}(\vartheta^0)^{\mathsf{T}}\right) \\ &=: \quad I_{n,1} + I_{n,2} + I_{n,3}. \end{aligned}$$

Note that the second term $I_{n,2}$ converges due to the ergodicity of $(\varepsilon_k^{(h)}(\vartheta^0))_{k\in\mathbb{N}}$, $\mathbb{E}(\varepsilon_k^{(h)}(\vartheta^0)\varepsilon_k^{(h)}(\vartheta^0)^{\mathsf{T}}) = V_{\vartheta^0}^{(h)}$ (see Lemma B.3(a,e)) and Birkhoff's Ergodic Theorem (see Bradley [12, 2.3 Ergodic Theorem]) so that

$$I_{n,2} \xrightarrow{a.s.} -\,\mathrm{tr}\left(\left(V_{\vartheta^0}^{(h)}\right)^{-1}\left(\partial_i^1 V_{\vartheta^0}^{(h)}\right)\left(V_{\vartheta^0}^{(h)}\right)^{-1}V_{\vartheta^0}^{(h)}\right) = -\,\mathrm{tr}\left(\left(V_{\vartheta^0}^{(h)}\right)^{-1}\partial_i^1 V_{\vartheta^0}^{(h)}\right).$$

Hence, $I_{n,1} + I_{n,2} \xrightarrow{a.s.} 0$. Thus, it only remains to show the convergence of the last term $I_{n,3}$. We obtain with Proposition A.1(a,c) and the continuous mapping theorem

$$
\begin{aligned}
&\frac{1}{n}\sum_{k=1}^{n}\big(\partial_i^1\varepsilon_k^{(h)}(\vartheta^0)\big)\varepsilon_k^{(h)}(\vartheta^0)^\mathsf{T}\\
&\quad= -\frac{1}{n}\sum_{k=1}^{n}[\big(\partial_i^1\Pi(\vartheta^0)\big)\,Y_{k-1}^{(h)}][\mathsf{k}(\mathsf{B},\vartheta^0)\Delta Y_k^{(h)}]^\mathsf{T}\\
&\qquad+\frac{1}{n}\sum_{k=1}^{n}[\big(\partial_i^1\Pi(\vartheta^0)\big)\,Y_{k-1}^{(h)}][\Pi(\vartheta^0)Y_{st,k-1}^{(h)}]^\mathsf{T}\\
&\qquad+\frac{1}{n}\sum_{k=1}^{n}[\big(\partial_i^1\mathsf{k}(\mathsf{B},\vartheta^0)\big)\,\Delta Y_k^{(h)}]\varepsilon_k^{(h)}(\vartheta^0)^\mathsf{T}\\
&\quad\xrightarrow{w} -\big(\partial_i^1\Pi(\vartheta^0)\big)\int_0^1 W_1(r)\mathrm{d}W_1(r)^\mathsf{T}\mathsf{k}(1,\vartheta^0)^\mathsf{T}\\
&\qquad-\sum_{j=1}^{\infty}\Gamma_{\partial_i^1\Pi(\vartheta^0)\Delta Y^{(h)},\mathsf{k}(\mathsf{B},\vartheta^0)\Delta Y^{(h)}}(j)\\
&\qquad+\big(\partial_i^1\Pi(\vartheta^0)\big)\int_0^1 W_1(r)\,\mathrm{d}W_2(r)^\mathsf{T}\Pi(\vartheta^0)^\mathsf{T}+\sum_{j=1}^{\infty}\Gamma_{\partial_i^1\Pi(\vartheta^0)\Delta Y^{(h)},\Pi(\vartheta^0)Y_{st}^{(h)}}(j)\\
&\qquad+\Gamma_{\partial_i^1\mathsf{k}(\mathsf{B},\vartheta^0)\Delta Y^{(h)},\varepsilon^{(h)}(\vartheta^0)}(0).
\end{aligned}\tag{4.3}
$$

Then, the continuous mapping theorem results in $I_{n,3}\xrightarrow{w}\mathcal{J}_1^{(i)}(\vartheta^0)$ which concludes the proof. □

4.1.2. Asymptotic behavior of the Hessian matrix

The second partial derivatives of the log-likelihood function $\mathcal{L}_n^{(h)}(\vartheta)$ are given by

$$
\begin{aligned}
&\partial_{i,j}\mathcal{L}_n^{(h)}(\vartheta)\\
&\quad=\operatorname{tr}\Big(\big(V_\vartheta^{(h)}\big)^{-1}\partial_{i,j}^2V_\vartheta^{(h)}-\big(V_\vartheta^{(h)}\big)^{-1}\big(\partial_iV_\vartheta^{(h)}\big)\big(V_\vartheta^{(h)}\big)^{-1}\big(\partial_jV_\vartheta^{(h)}\big)\Big)\\
&\qquad-\frac{1}{n}\sum_{k=1}^{n}\operatorname{tr}\Big(\big(V_\vartheta^{(h)}\big)^{-1}\varepsilon_k^{(h)}(\vartheta)\varepsilon_k^{(h)}(\vartheta)^\mathsf{T}\big(V_\vartheta^{(h)}\big)^{-1}\partial_{i,j}^2V_\vartheta^{(h)}\Big)\\
&\qquad+\frac{1}{n}\sum_{k=1}^{n}\operatorname{tr}\Big(\big(V_\vartheta^{(h)}\big)^{-1}\big(\partial_jV_\vartheta^{(h)}\big)\big(V^{(h)}\big)^{-1}\varepsilon_k^{(h)}(\vartheta)\varepsilon_k^{(h)}(\vartheta)^\mathsf{T}\big(V_\vartheta^{(h)}\big)^{-1}\partial_iV_\vartheta^{(h)}\Big)\\
&\qquad+\frac{1}{n}\sum_{k=1}^{n}\operatorname{tr}\Big(\big(V_\vartheta^{(h)}\big)^{-1}\varepsilon_k^{(h)}(\vartheta)\varepsilon_k^{(h)}(\vartheta)^\mathsf{T}\big(V_\vartheta^{(h)}\big)^{-1}\big(\partial_jV_\vartheta^{(h)}\big)\big(V_\vartheta^{(h)}\big)^{-1}\partial_iV_\vartheta^{(h)}\Big)\\
&\qquad-\frac{1}{n}\sum_{k=1}^{n}\operatorname{tr}\Big(\big(V_\vartheta^{(h)}\big)^{-1}\big(\partial_j\varepsilon_k^{(h)}(\vartheta)\varepsilon_k^{(h)}(\vartheta)^\mathsf{T}\big)\big(V_\vartheta^{(h)}\big)^{-1}\partial_iV_\vartheta^{(h)}\Big)
\end{aligned}
$$

$$
\begin{aligned}
&+\frac{2}{n}\sum_{k=1}^{n}\left(\partial_{i,j}\varepsilon_k^{(h)}(\vartheta)^\mathsf{T}\right)\left(V_\vartheta^{(h)}\right)^{-1}\varepsilon_k^{(h)}(\vartheta)\\
&-\frac{2}{n}\sum_{k=1}^{n}\operatorname{tr}\left(\left(V_\vartheta^{(h)}\right)^{-1}\varepsilon_k^{(h)}(\vartheta)\left(\partial_i\varepsilon_k^{(h)}(\vartheta)^\mathsf{T}\right)\left(V_\vartheta^{(h)}\right)^{-1}\partial_j V_\vartheta^{(h)}\right)\\
&+\frac{2}{n}\sum_{k=1}^{n}\left(\partial_i\varepsilon_k^{(h)}(\vartheta)^\mathsf{T}\right)\left(V_\vartheta^{(h)}\right)^{-1}\left(\partial_j\varepsilon_k^{(h)}(\vartheta)\right)\\
=:&\sum_{j=1}^{8} I_{n,j}.
\end{aligned} \tag{4.4}
$$

Since the Hessian matrix should be asymptotically positive definite we need an additional assumption.

Assumption E. *The $((d-c)c \times s_1)$-dimensional gradient matrix $\nabla_{\vartheta_1}\left(C_{1,\vartheta_1}^{\perp\mathsf{T}}C_1\right)\Big|_{\vartheta_1=\vartheta_1^0}$ is of full column rank s_1.*

The asymptotic distribution of the Hessian matrix is given in the next proposition.

Proposition 4.2. *Let Assumption E additionally hold. Define the $(s_1 \times s_1)$-dimensional random matrix $Z_1(\vartheta^0)$ as*

$$
[Z_1(\vartheta^0)]_{i,j} := 2\cdot\operatorname{tr}\left(\left(V_{\vartheta^0}^{(h)}\right)^{-1}\partial_i^1\Pi(\vartheta^0)\int_0^1 W_1(r)W_1(r)^\mathsf{T}\,\mathrm{d}r\,\left(\partial_j^1\Pi(\vartheta^0)\right)^\mathsf{T}\right)
$$

for $i,j=1,\ldots,s_1$. Then, $Z_1(\vartheta^0)$ is almost surely positive definite and

$$
n^{-1}\nabla_{\vartheta_1}^2\mathcal{L}_n^{(h)}(\vartheta^0)\xrightarrow{w} Z_1(\vartheta^0).
$$

Proof. First, we prove the asymptotic behavior of the score vector and then, in the next step, that the limit is almost surely positive definite.

Step 1: The first term $\frac{1}{n}I_{n,1}$ in (4.4) converges to zero due to the additional normalizing rate of n^{-1}. Due to Proposition A.1 (a,c) we have for $j=2,\ldots,7$ that $I_{n,j}=\mathcal{O}_p(1)$ (see exemplarily (4.3) for $I_{n,5}$) and hence, $\frac{1}{n}\sum_{j=2}^{7}I_{n,j}$ converges in probability to zero. To summarize,

$$
\begin{aligned}
n^{-1}\partial_{i,j}^1\mathcal{L}_n^{(h)}(\vartheta^0) &= \frac{1}{n}I_{n,8}+o_p(1)\\
&= 2\operatorname{tr}\left(\left(V_{\vartheta^0}^{(h)}\right)^{-1}\frac{1}{n^2}\sum_{k=1}^{n}\partial_i^1\varepsilon_k^{(h)}(\vartheta^0)\partial_j^1\varepsilon_k^{(h)}(\vartheta^0)^\mathsf{T}\right)+o_p(1).
\end{aligned}
$$

Due to Lemma B.2 and Proposition A.1 (a,c) we receive

$$
n^{-1}\partial_{i,j}^1\mathcal{L}_n^{(h)}(\vartheta^0) = 2\operatorname{tr}\left(\left(V_{\vartheta^0}^{(h)}\right)^{-1}\frac{1}{n^2}\sum_{k=1}^{n}\partial_i^1\Pi(\vartheta^0)Y_{k-1}^{(h)}Y_{k-1}^{(h)\mathsf{T}}\partial_j^1\Pi(\vartheta^0)^\mathsf{T}\right)+o_p(1).
$$

Then, Proposition A.1(b) and the continuous mapping theorem result in

$$n^{-1}\partial^1_{i,j}\mathcal{L}^{(h)}_n(\vartheta^0) \xrightarrow{w} 2\,\mathrm{tr}\left(\left(V^{(h)}_{\vartheta^0}\right)^{-1}\partial^1_i\Pi(\vartheta^0)\int_0^1 W_1(r)W_1(r)^\mathsf{T}\,\mathrm{d}r\,\left(\partial^1_j\Pi(\vartheta^0)\right)^\mathsf{T}\right).$$

In particular, we have also the joint convergence of the partial derivatives.

Step 2: Let $W_1 = C_1B_1W_3$ and define $M := B_1\int_0^1 W_3(r)W_3(r)^\mathsf{T}\,\mathrm{d}rB_1^\mathsf{T}$, which is a $\mathbb{P}$-a.s. positive definite $c\times c$ matrix. We apply the Cholesky decomposition $M = M_*M_*^\mathsf{T}$. By using properties of the vec operator and the Kronecker product (see Bernstein [9, Chapter 7.1]) we have

$$\begin{aligned}
&[Z_1(\vartheta^0)]_{i,j} \qquad\qquad (4.5)\\
&= 2\,\mathrm{tr}\left(\left(V^{(h)}_{\vartheta^0}\right)^{-\frac12}\partial^1_i\Pi(\vartheta^0)C_1M\left(\left(V^{(h)}_{\vartheta^0}\right)^{-\frac12}\partial^1_j\Pi(\vartheta^0)C_1\right)^\mathsf{T}\right)\\
&= 2\mathrm{vec}\Big(\left(V^{(h)}_{\vartheta^0}\right)^{-\frac12}\alpha(\vartheta^0)\partial^1_iC^{\perp\mathsf{T}}_{1,\vartheta^0_1}C_1M_*\Big)^\mathsf{T}\mathrm{vec}\Big(\left(V^{(h)}_{\vartheta^0}\right)^{-\frac12}\alpha(\vartheta^0)\partial^1_jC^{\perp\mathsf{T}}_{1,\vartheta^0_1}C_1M_*\Big)\\
&= 2\mathrm{vec}\left(\partial^1_iC^{\perp\mathsf{T}}_{1,\vartheta^0_1}C_1\right)^\mathsf{T}\left(M\otimes\left(\alpha(\vartheta^0)^\mathsf{T}\left(V^{(h)}_{\vartheta^0}\right)^{-1}\alpha(\vartheta^0)\right)\right)\mathrm{vec}\left(\partial^1_jC^{\perp\mathsf{T}}_{1,\vartheta^0_1}C_1\right).
\end{aligned}$$

Furthermore,

$$\mathrm{rank}\left(M\otimes\left(\alpha(\vartheta^0)^\mathsf{T}\left(V^{(h)}_{\vartheta^0}\right)^{-1}\alpha(\vartheta^0)\right)\right) = \mathrm{rank}(M)\cdot\mathrm{rank}\left(\alpha(\vartheta^0)^\mathsf{T}\left(V^{(h)}_{\vartheta^0}\right)^{-1}\alpha(\vartheta^0)\right)$$

due to Bernstein [9, Fact 7.4.23] and thus, $M\otimes\left(\alpha(\vartheta^0)^\mathsf{T}\left(V^{(h)}_{\vartheta^0}\right)^{-1}\alpha(\vartheta^0)\right)$ has full rank $c\cdot(d-c)$ a.s. Now, if we consider the Hessian matrix $Z_1(\vartheta^0)$, we have

$$\begin{aligned}
&Z_1(\vartheta^0) =\\
&2\left[\nabla_{\vartheta_1}\left(C^{\perp\mathsf{T}}_{1,\vartheta_1}C_1\right)^\mathsf{T}\right]_{\vartheta_1=\vartheta^0_1}M\otimes\left(\alpha(\vartheta^0)^\mathsf{T}\left(V^{(h)}_{\vartheta^0}\right)^{-1}\alpha(\vartheta^0)\right)\left[\nabla_{\vartheta_1}\left(C^{\perp\mathsf{T}}_{1,\vartheta_1}C_1\right)\right]_{\vartheta_1=\vartheta^0_1}.
\end{aligned}$$

Due to Assumption E the $((d-c)c\times s_1)$-dimensional matrix $\nabla_{\vartheta_1}\left(C^{\perp\mathsf{T}}_{1,\vartheta^0_1}C_1\right)$ is of full column rank and hence, the product has full rank s_1. Therefore, we have the positive definiteness almost surely. □

4.1.3. Asymptotic mixed normality of the long-run QML estimator

We are able now to show the weak convergence of the long-run QML estimator and thus, we have one main result.

Theorem 4.3. *Let Assumption E additionally hold. Then, we have as $n\to\infty$*

$$n(\widehat{\vartheta}_{n,1}-\vartheta^0_1)\xrightarrow{w} -Z_1(\vartheta^0)^{-1}\cdot\mathcal{J}_1(\vartheta^0),$$

where $\mathcal{J}_1(\vartheta^0)$ is defined as in Proposition 4.1 and $Z_1(\vartheta^0)$ as in Proposition 4.2, respectively.

Proof. From (4.1) we know that

$$0_{s_1} = \nabla_{\vartheta_1}\widehat{\mathcal{L}}_n^{(h)}(\vartheta_1^0,\widehat{\vartheta}_{n,2}) + n^{-1}\nabla_{\vartheta_1}^2\underline{\widehat{\mathcal{L}}}_n^{(h)}(\underline{\vartheta}_{n,1},\widehat{\vartheta}_{n,2})n(\widehat{\vartheta}_{n,1}-\vartheta_1^0). \tag{4.6}$$

In Proposition 4.1 we already derived the asymptotic behavior of the score vector $\nabla_{\vartheta_1}\mathcal{L}_n^{(h)}(\vartheta_1^0,\vartheta_2^0)$ and in Proposition 4.2 the asymptotic behavior of the Hessian matrix $n^{-1}\nabla_{\vartheta_1}^2\mathcal{L}_n^{(h)}(\vartheta_1^0,\vartheta_2^0)$. However, for the proof of Theorem 4.3 we require now the asymptotic behavior of $\nabla_{\vartheta_1}\mathcal{L}_n^{(h)}(\vartheta_1^0,\widehat{\vartheta}_{n,2})$ and $n^{-1}\nabla_{\vartheta_1}^2\underline{\mathcal{L}}_n^{(h)}(\underline{\vartheta}_{n,1},\widehat{\vartheta}_{n,2})$. Therefore, we use a local stochastic equicontinuity condition on the family $\nabla_{\vartheta_1}\mathcal{L}_n^{(h)}(\vartheta_1^0,\cdot)$ in ϑ_2^0 $(n\in\mathbb{N})$ and on the family $n^{-1}\nabla_{\vartheta_1}^2\mathcal{L}_n^{(h)}(\cdot)$ in ϑ^0 $(n\in\mathbb{N})$.

Lemma 4.4. *For every $\tau>0$ and every $\eta>0$, there exist an integer $n(\tau,\eta)$ and real numbers $\delta_1,\delta_2>0$ such that for $\frac{1}{2}<\gamma<1$ and $n\geq n(\tau,\eta)$:*

(a) $\mathbb{P}\left(\sup_{\vartheta_2\in\mathcal{B}(\vartheta_2^0,\delta_2)}\|\nabla_{\vartheta_1}\mathcal{L}_n^{(h)}(\vartheta_1^0,\vartheta_2)-\nabla_{\vartheta_1}\mathcal{L}_n^{(h)}(\vartheta_1^0,\vartheta_2^0)\|>\tau\right)\leq\eta,$

(b) $\mathbb{P}\left(\sup_{\vartheta\in N_{n,\gamma}(\vartheta_1^0,\delta_1)\times\mathcal{B}(\vartheta_2^0,\delta_2)}\|n^{-1}\nabla_{\vartheta_1}^2\mathcal{L}_n^{(h)}(\vartheta)-n^{-1}\nabla_{\vartheta_1}^2\mathcal{L}_n^{(h)}(\vartheta^0)\|>\tau\right)\leq\eta.$

The stochastic equicontinuity conditions SE and SE_o in Saikkonen [47] are global conditions whereas Lemma 4.3 is weaker and presents only a local stochastic equicontinuity condition for the standardized score in ϑ_2^0 and for the standardized Hessian matrix in ϑ^0.

Proof of Lemma 4.4.
(a) Note that on the one hand, $\nabla_{\vartheta_1}\mathcal{L}_{n,1,1}^{(h)}(\vartheta_1^0,\vartheta_2)=0$ since $\varepsilon_{n,1,1}^{(h)}(\vartheta_1^0,\vartheta_2)=0$ and on the other hand, $\nabla_{\vartheta_1}\mathcal{L}_{n,2}^{(h)}(\vartheta_2)=0$. Hence,

$$\begin{aligned}&\sup_{\vartheta_2\in\mathcal{B}(\vartheta_2^0,\delta_2)}\|\nabla_{\vartheta_1}\mathcal{L}_n^{(h)}(\vartheta_1^0,\vartheta_2)-\nabla_{\vartheta_1}\mathcal{L}_n^{(h)}(\vartheta_1^0,\vartheta_2^0)\|\\&\quad=\sup_{\vartheta_2\in\mathcal{B}(\vartheta_2^0,\delta_2)}\|\nabla_{\vartheta_1}\mathcal{L}_{n,1,2}^{(h)}(\vartheta_1^0,\vartheta_2)-\nabla_{\vartheta_1}\mathcal{L}_{n,1,2}^{(h)}(\vartheta_1^0,\vartheta_2^0)\|.\end{aligned}$$

We can conclude with similar calculations as in Lemma 3.3 applying (A.4) and (A.6) that

$$\begin{aligned}&\sup_{\vartheta_2\in\mathcal{B}(\vartheta_2^0,\delta_2)}\|\nabla_{\vartheta_1}\mathcal{L}_{n,1,2}^{(h)}(\vartheta_1^0,\vartheta_2)-\nabla_{\vartheta_1}\mathcal{L}_{n,1,2}^{(h)}(\vartheta_1^0,\vartheta_2^0)\|\\&\quad\leq\sup_{\vartheta_2\in\mathcal{B}(\vartheta_2^0,\delta_2)}\mathfrak{C}\|\vartheta_2-\vartheta_2^0\|U_n\leq\mathfrak{C}\delta_2U_n.\end{aligned}$$

Since $U_n=\mathcal{O}_p(1)$ due to Lemma 3.3 we obtain the statement.
(b) Due to $\nabla_{\vartheta_1}^2\mathcal{L}_{n,2}^{(h)}(\vartheta_2)=0$ we have

$$\begin{aligned}&\sup_{\vartheta\in N_{n,\gamma}(\vartheta_1^0,\delta_1)\times\mathcal{B}(\vartheta_2^0,\delta_2)}\|n^{-1}\nabla_{\vartheta_1}^2\mathcal{L}_n^{(h)}(\vartheta)-n^{-1}\nabla_{\vartheta_1}^2\mathcal{L}_n^{(h)}(\vartheta^0)\|\\&\quad\leq\sup_{\vartheta\in N_{n,\gamma}(\vartheta_1^0,\delta_1)\times\mathcal{B}(\vartheta_2^0,\delta_2)}\|n^{-1}\nabla_{\vartheta_1}^2\mathcal{L}_{n,1,1}^{(h)}(\vartheta)-n^{-1}\nabla_{\vartheta_1}^2\mathcal{L}_{n,1,1}^{(h)}(\vartheta^0)\|\end{aligned}$$

$$+ \sup_{\vartheta \in N_{n,\gamma}(\vartheta_1^0,\delta_1)\times\mathcal{B}(\vartheta_2^0,\delta_2)} \|n^{-1}\nabla^2_{\vartheta_1}\mathcal{L}^{(h)}_{n,1,2}(\vartheta) - n^{-1}\nabla^2_{\vartheta_1}\mathcal{L}^{(h)}_{n,1,2}(\vartheta^0)\|.$$

Then, the first term is bounded by (A.2) and the second term by (A.4) and (A.6), respectively. Hence,

$$\begin{aligned}
&\leq \sup_{\vartheta \in N_{n,\gamma}(\vartheta_1^0,\delta_1)\times\mathcal{B}(\vartheta_2^0,\delta_2)} \left(\mathfrak{C}\|\vartheta-\vartheta^0\| \left\|\frac{1}{n^2}\sum_{k=1}^n L^{(h)}_{k-1}[L^{(h)}_{k-1}]^\mathsf{T}\right\| + \frac{1}{n}\mathfrak{C}\|\vartheta-\vartheta^0\|U_n\right) \\
&\leq \mathfrak{C}\delta_2 \left\|\frac{1}{n^2}\sum_{k=1}^n L^{(h)}_{k-1}[L^{(h)}_{k-1}]^\mathsf{T}\right\| + \frac{1}{n}\mathfrak{C}\delta_2 U_n.
\end{aligned}$$

Since $U_n = \mathcal{O}_p(1)$ due to Lemma 3.3 and $\frac{1}{n^2}\sum_{k=1}^n L^{(h)}_{k-1}[L^{(h)}_{k-1}]^\mathsf{T} = \mathcal{O}_p(1)$ due to Proposition A.1(b), statement (b) follows. □

The weak convergence of $\nabla_{\vartheta_1}\widehat{\mathcal{L}}^{(h)}_n(\vartheta_1^0,\widehat{\vartheta}_{n,2})$ to $\mathcal{J}_1(\vartheta^0)$ follows then by Proposition 2.5, Proposition 4.1 and Lemma 4.4(a). Due to Proposition 2.5, Proposition 4.2 and Lemma 4.4(b) we have that $n^{-1}\nabla^2_{\vartheta_1}\widehat{\underline{\mathcal{L}}}^{(h)}_n(\underline{\vartheta}_{n,1},\widehat{\vartheta}_{n,2})$ converges weakly to the random matrix $Z_1(\vartheta^0)$. In particular, Proposition A.1 also guarantees the joint convergence of both terms. Finally, the almost sure positive definiteness of $Z_1(\vartheta^0)$ allows us to take the inverse and reorder (4.1) so that

$$\begin{aligned}
n(\widehat{\vartheta}_{n,1}-\vartheta_1^0) &= -\left(n^{-1}\nabla^2_{\vartheta_1}\widehat{\underline{\mathcal{L}}}^{(h)}_n(\underline{\vartheta}_{n,1},\widehat{\vartheta}_{n,2})\right)^{-1}\nabla_{\vartheta_1}\widehat{\mathcal{L}}^{(h)}_n(\vartheta_1^0,\widehat{\vartheta}_{n,2}) \\
&\xrightarrow{w} -Z_1(\vartheta^0)^{-1}\cdot\mathcal{J}_1(\vartheta^0).
\end{aligned}$$
□

4.2. Asymptotic distribution of the short-run parameter estimator

Lastly, we derive the asymptotic normality of the short-run QML estimator $\widehat{\vartheta}_{n,2}$ which we prove by using a Taylor-expansion of the QML-function similarly as in Section 4.1. Before we state the proof we want to derive some mixing property of the process $(Y^{(h)}_{st,k},\Delta Y^{(h)}_k)_{k\in\mathbb{Z}}$ because this will be used throughout this section.

Lemma 4.5. *The process $(Y^{(h)}_{st,k},\Delta Y^{(h)}_k)_{k\in\mathbb{Z}}$ is strongly mixing with mixing coefficients $\alpha_{\Delta Y^{(h)},Y^{(h)}_{st}}(l)\leq \mathfrak{C}\rho^l$ for some $0<\rho<1$. In particular, for any $\delta>0$,*

$$\sum_{l=1}^\infty \alpha_{\Delta Y^{(h)},Y^{(h)}_{st}}(l)^{\frac{\delta}{2+\delta}} < \infty.$$

Proof. Due to (2.2) the process $Y^{(h)}_{st}$ has the state space representation

$$Y^{(h)}_{st,k} = C_2 X^{(h)}_{st,k} \quad \text{with} \quad X^{(h)}_{st,k} = \mathrm{e}^{A_2 h}X^{(h)}_{st,k-1} + \int_{(k-1)h}^{kh}\mathrm{e}^{A_2(kh-u)}B_2\,dL_u$$

for $k \in \mathbb{N}$. Masuda [39, Theorem 4.3] proved that $(X_{st,k}^{(h)})_{k\in\mathbb{N}}$ is β-mixing with an exponentially rate since $\mathbb{E}\|X_{st,k}^{(h)}\|^2 < \infty$. Having $\mathbb{E}\|\Delta L_k^{(h)}\|^2 < \infty$ in mind as well we can conclude on the same way as in Masuda [39, Theorem 4.3] that the Markov process

$$\begin{aligned}\begin{pmatrix} \Delta L_k^{(h)} \\ X_{st,k}^{(h)} \end{pmatrix} &= \begin{pmatrix} 0_{m\times m} & 0_{m\times(N-c)} \\ 0_{(N-c)\times m} & \mathrm{e}^{A_2 h} \end{pmatrix} \begin{pmatrix} \Delta L_{k-1}^{(h)} \\ X_{st,k-1}^{(h)} \end{pmatrix} \\ &\quad + \int_{(k-1)h}^{kh} \begin{pmatrix} I_m \\ \mathrm{e}^{A_2(kh-u)} B_2 \end{pmatrix} dL_u\end{aligned}$$

is β-mixing with mixing coefficient $\beta_{\Delta L^{(h)},X^{(h)}}(l) \leq \mathfrak{C}\rho_1^l$ for some $0 < \rho_1 < 1$. Hence, it is as well α-mixing with mixing coefficient

$$\alpha_{\Delta L^{(h)},X^{(h)}}(l) \leq \beta_{\Delta L^{(h)},X^{(h)}}(l) \leq \mathfrak{C}\rho_1^l.$$

Finally, it is obvious of the definition of α-mixing that

$$\begin{pmatrix} \Delta Y_k^{(h)} \\ Y_{st,k}^{(h)} \end{pmatrix} = \begin{pmatrix} C_1 B_1 & C_2 & -C_2 \\ 0_{d\times m} & C_2 & 0_{d\times N-c} \end{pmatrix} \begin{pmatrix} \Delta L_k^{(h)} \\ X_{st,k}^{(h)} \\ X_{st,k-1}^{(h)} \end{pmatrix}$$

is α-mixing with $\alpha_{\Delta Y^{(h)},Y_{st}^{(h)}}(l) \leq \alpha_{\Delta L^{(h)},X^{(h)}}(l-1) \leq \mathfrak{C}\rho_1^{l-1}$. □

4.2.1. Asymptotic behavior of the score vector

First, we prove that the partial derivatives have finite variance.

Lemma 4.6. *$\mathbb{E}|\partial_i^{st}\mathcal{L}_n^{(h)}(\vartheta^0)|^2 < \infty$ for any $n \in \mathbb{N}$ and $i = 1,\ldots,s_2$.*

Proof. We have due to Lemma B.3(b) and the Cauchy-Schwarz inequality

$$\begin{aligned}&\mathbb{E}\Big| -\mathrm{tr}\left(\left(V_{\vartheta^0}^{(h)}\right)^{-1}\varepsilon_k^{(h)}(\vartheta^0)\varepsilon_k^{(h)}(\vartheta^0)^\mathsf{T}\left(V_{\vartheta^0}^{(h)}\right)^{-1}\partial_i^{st}V_{\vartheta^0}^{(h)}\right) \\ &\qquad 2\cdot\left(\partial_i^{st}\varepsilon_k^{(h)}(\vartheta^0)^\mathsf{T}\right)\left(V_{\vartheta^0}^{(h)}\right)^{-1}\varepsilon_k^{(h)}(\vartheta^0)\Big|^2 \\ &\leq \mathfrak{C}\cdot\mathbb{E}\|\varepsilon_k^{(h)}(\vartheta^0)\|^4 + \mathfrak{C}\cdot\left(\mathbb{E}\|\varepsilon_k^{(h)}(\vartheta^0)\|^4\mathbb{E}\|\partial_i^{st}\varepsilon_k^{(h)}(\vartheta^0)\|^4\right)^{\frac{1}{2}} < \infty,\end{aligned}$$

so that the statement follows with (4.2). □

Now we can prove the convergence of the covariance matrix of the score vector where we plug in the true parameter.

Lemma 4.7. *Define $\ell_{k,2}^{(h)}(\vartheta_2) := \varepsilon_{k,2}^{(h)}(\vartheta_2)^\mathsf{T}\left(V_{\vartheta_1^0,\vartheta_2}^{(h)}\right)^{-1}\varepsilon_{k,2}^{(h)}(\vartheta_2)$ for $\vartheta_2 \in \Theta_2$. Then,*

$$\begin{aligned}\lim_{n\to\infty}\mathrm{Var}\left(\nabla_{\vartheta_2}\mathcal{L}_n^{(h)}(\vartheta^0)\right) &= \Big[\sum_{l\in\mathbb{Z}}\mathrm{Cov}\left(\partial_i^{st}\ell_{1,2}^{(h)}(\vartheta_2^0), \partial_j^{st}\ell_{1+l,2}^{(h)}(\vartheta_2^0)\right)\Big]_{1\leq i,j\leq s_2} \\ &=: I(\vartheta_2^0). \qquad (4.7)\end{aligned}$$

Proof. We can derive the result in a similar way as in Schlemm and Stelzer [51, Lemma 2.14]. Hence, we only sketch the proof to show the differences. A detailed proof can be found in Scholz [52, Section 5.9]. It is sufficient to show that for all $i, j = 1, \ldots, s_2$ the sequence

$$\begin{aligned} [\mathrm{Var}\left(\nabla_{\vartheta_2}\mathcal{L}_n^{(h)}(\vartheta^0)\right)]_{i,j} &= n^{-1}\sum_{k_1=1}^{n}\sum_{k_2=1}^{n}\mathrm{Cov}\left(\partial_i^{st}\ell_{k_1,2}^{(h)}(\vartheta_2^0), \partial_j^{st}\ell_{k_2,2}^{(h)}(\vartheta_2^0)\right) \\ &=: I_n^{(i,j)}(\vartheta_2^0) \end{aligned} \tag{4.8}$$

converges as $n \to \infty$. By the representation of the partial derivatives in (4.2) and (B.1) the sequence

$$\begin{aligned} \partial_i^{st}\ell_{k,2}^{(h)}(\vartheta_2^0) &= -\operatorname{tr}\left(\left(V_{\vartheta^0}^{(h)}\right)^{-1}\varepsilon_{k,2}^{(h)}(\vartheta_2^0)\varepsilon_{k,2}^{(h)}(\vartheta_2^0)^{\mathsf{T}}\left(V_{\vartheta^0}^{(h)}\right)^{-1}\partial_i^{st}V_{\vartheta^0}^{(h)}\right) \\ &\quad+\left(\partial_i^{st}\varepsilon_{k,2}^{(h)}(\vartheta_2^0)^{\mathsf{T}}\right)\left(V_{\vartheta^0}^{(h)}\right)^{-1}\varepsilon_{k,2}^{(h)}(\vartheta_2^0) \\ &= -\operatorname{tr}\left(\left(V_{\vartheta^0}^{(h)}\right)^{-1}\varepsilon_{k}^{(h)}(\vartheta^0)\varepsilon_{k}^{(h)}(\vartheta^0)^{\mathsf{T}}\left(V_{\vartheta^0}^{(h)}\right)^{-1}\partial_i^{st}V_{\vartheta^0}^{(h)}\right) \\ &\quad+\left(\partial_i^{st}\varepsilon_{k}^{(h)}(\vartheta^0)^{\mathsf{T}}\right)\left(V_{\vartheta^0}^{(h)}\right)^{-1}\varepsilon_{k}^{(h)}(\vartheta^0) \end{aligned}$$

is stationary and the covariance in (4.8) depends only on the difference $l = k_1 - k_2$. If we can show that

$$\sum_{l\in\mathbb{Z}}\left|\mathrm{Cov}\left(\partial_i^{st}\ell_{1,2}^{(h)}(\vartheta_2^0), \partial_j^{st}\ell_{1+l,2}^{(h)}(\vartheta_2^0)\right)\right| < \infty, \tag{4.9}$$

then the Dominated Convergence Theorem implies

$$\begin{aligned} I_n^{(i,j)}(\vartheta_2^0) &= n^{-1}\sum_{l=-n}^{n}(n-|l|)\,\mathrm{Cov}\left(\partial_i^{st}\ell_{1,2}^{(h)}(\vartheta_2^0), \partial_j^{st}\ell_{1+l,2}^{(h)}(\vartheta_2^0)\right) \\ &\xrightarrow{n\to\infty} \sum_{l\in\mathbb{Z}}\mathrm{Cov}\left(\partial_i^{st}\ell_{1,2}^{(h)}(\vartheta_2^0), \partial_j^{st}\ell_{1+l,2}^{(h)}(\vartheta_2^0)\right). \end{aligned}$$

Due to Lemma 4.5 and the uniformly exponentially bound of $(\mathsf{k}_j(\vartheta))$ and $(\partial_i^{st}\mathsf{k}_j(\vartheta))$ finding the dominant goes in the same vein as in the proof of Schlemm and Stelzer [51, Lemma 2.14]. ☐

In the following, we derive the convergence of the score vector with respect to the short-run parameters by a truncation argument.

Proposition 4.8. *For the gradient with respect to the short-run parameters the asymptotic behavior*

$$\sqrt{n}\cdot\nabla_{\vartheta_2}\mathcal{L}_n^{(h)}(\vartheta^0) \xrightarrow{w} \mathcal{N}(0, I(\vartheta_2^0))$$

holds, where $I(\vartheta_2^0)$ *is the asymptotic covariance matrix given in* (4.7).

Proof. First, we realize that representation (4.2) and Lemma B.3(c,d) result in $\mathbb{E}\big(\nabla_{\vartheta_2}\mathcal{L}_n^{(h)}(\vartheta^0)\big) = 0_{s_2}$. Due to (B.1) we can rewrite (4.2) for $\mathsf{M} \in \mathbb{N}$ as

$$\partial_i^{st}\mathcal{L}_n^{(h)}(\vartheta^0) = \tfrac{1}{n}\textstyle\sum_{k=1}^n \left(Y_{\mathsf{M},k}^{(i)} - \mathbb{E}Y_{\mathsf{M},k}^{(i)}\right) + \tfrac{1}{n}\textstyle\sum_{k=1}^n \left(Z_{\mathsf{M},k}^{(i)} - \mathbb{E}Z_{\mathsf{M},k}^{(i)}\right), \quad (4.10)$$

where

$$\begin{aligned}
Y_{\mathsf{M},k}^{(i)} :=\ & \operatorname{tr}\left(\big(V_{\vartheta^0}^{(h)}\big)^{-1}\partial_i^{st}V_{\vartheta^0}^{(h)}\right) \\
& - \operatorname{tr}\left(\big(V_{\vartheta^0}^{(h)}\big)^{-1}\Pi(\vartheta^0)Y_{st,k-1}^{(h)}Y_{st,k-1}^{(h)\mathsf{T}}\Pi(\vartheta^0)^{\mathsf{T}}\big(V_{\vartheta^0}^{(h)}\big)^{-1}\partial_i^{st}V_{\vartheta^0}^{(h)}\right) \\
& + \sum_{\iota_1=0}^{\mathsf{M}} \operatorname{tr}\left(\big(V_{\vartheta^0}^{(h)}\big)^{-1}\mathsf{k}_{\iota_1}(\vartheta^0)\Delta Y_{k-\iota_1}^{(h)}Y_{st,k-1}^{(h)\mathsf{T}}\Pi(\vartheta^0)^{\mathsf{T}}\big(V_{\vartheta^0}^{(h)}\big)^{-1}\partial_i^{st}V_{\vartheta^0}^{(h)}\right) \\
& + \sum_{\iota_2=0}^{\mathsf{M}} \operatorname{tr}\left(\big(V_{\vartheta^0}^{(h)}\big)^{-1}\Pi(\vartheta^0)Y_{st,k-1}^{(h)}\Delta Y_{k-\iota_2}^{(h)\mathsf{T}}\mathsf{k}_{\iota_2}(\vartheta^0)^{\mathsf{T}}\big(V_{\vartheta^0}^{(h)}\big)^{-1}\partial_i^{st}V_{\vartheta^0}^{(h)}\right) \\
& - \sum_{\iota_1,\iota_2=0}^{\mathsf{M}} \operatorname{tr}\left(\big(V_{\vartheta^0}^{(h)}\big)^{-1}\mathsf{k}_{\iota_1}(\vartheta^0)\Delta Y_{k-\iota_1}^{(h)}\Delta Y_{k-\iota_2}^{(h)\mathsf{T}}\mathsf{k}_{\iota_2}(\vartheta^0)^{\mathsf{T}}\big(V_{\vartheta^0}^{(h)}\big)^{-1}\partial_i^{st}V_{\vartheta^0}^{(h)}\right) \\
& + 2\cdot\operatorname{tr}\left(\big(\partial_i^{st}\Pi(\vartheta^0)Y_{st,k-1}^{(h)}\big)\big(V_{\vartheta^0}^{(h)}\big)^{-1}Y_{st,k-1}^{(h)\mathsf{T}}\Pi(\vartheta^0)^{\mathsf{T}}\right) \\
& - 2\cdot\sum_{\iota_1=0}^{\mathsf{M}} \operatorname{tr}\left(\big(\partial_i^{st}\mathsf{k}_{\iota_1}(\vartheta^0)\Delta Y_{k-\iota_1}^{(h)}\big)\big(V_{\vartheta^0}^{(h)}\big)^{-1}Y_{st,k-1}^{(h)\mathsf{T}}\Pi(\vartheta^0)^{\mathsf{T}}\right) \\
& - 2\cdot\sum_{\iota_2=0}^{\mathsf{M}} \operatorname{tr}\left(\big(\partial_i^{st}\Pi(\vartheta^0)Y_{st,k-1}^{(h)}\big)\big(V_{\vartheta^0}^{(h)}\big)^{-1}\Delta Y_{k-\iota_2}^{(h)\mathsf{T}}\mathsf{k}_{\iota_2}(\vartheta^0)^{\mathsf{T}}\right) \\
& + 2\cdot\sum_{\iota_1,\iota_2=0}^{\mathsf{M}} \operatorname{tr}\left(\big(\partial_i^{st}\mathsf{k}_{\iota_1}(\vartheta^0)\Delta Y_{k-\iota_1}^{(h)}\big)\big(V_{\vartheta^0}^{(h)}\big)^{-1}\Delta Y_{k-\iota_2}^{(h)\mathsf{T}}\mathsf{k}_{\iota_2}(\vartheta^0)^{\mathsf{T}}\right), \\
Z_{\mathsf{M},k}^{(i)} :=\ & V_{\mathsf{M},k}^{(i)} + U_{\mathsf{M},k}^{(i)},
\end{aligned}$$

and

$$\begin{aligned}
V_{\mathsf{M},k}^{(i)} := & \sum_{\iota_1=\mathsf{M}+1}^{\infty} \operatorname{tr}\left(\big(V_{\vartheta^0}^{(h)}\big)^{-1}\mathsf{k}_{\iota_1}(\vartheta^0)\Delta Y_{k-\iota_1}^{(h)}Y_{st,k-1}^{(h)\mathsf{T}}\Pi(\vartheta^0)^{\mathsf{T}}\big(V_{\vartheta^0}^{(h)}\big)^{-1}\partial_i^{st}V_{\vartheta^0}^{(h)}\right) \\
& - \sum_{\iota_1=\mathsf{M}+1}^{\infty}\sum_{\iota_2=0}^{\infty} \operatorname{tr}\left(\big(V_{\vartheta^0}^{(h)}\big)^{-1}\mathsf{k}_{\iota_1}(\vartheta^0)\Delta Y_{k-\iota_1}^{(h)}\Delta Y_{k-\iota_2}^{(h)\mathsf{T}}\mathsf{k}_{\iota_2}(\vartheta^0)^{\mathsf{T}}\big(V_{\vartheta^0}^{(h)}\big)^{-1}\partial_i^{st}V_{\vartheta^0}^{(h)}\right) \\
& - 2\cdot\sum_{\iota_1=\mathsf{M}+1}^{\infty} \operatorname{tr}\left(\big(\partial_i^{st}\mathsf{k}_{\iota_1}(\vartheta^0)\Delta Y_{k-\iota_1}^{(h)}\big)\big(V_{\vartheta^0}^{(h)}\big)^{-1}Y_{st,k-1}^{(h)\mathsf{T}}\Pi(\vartheta^0)^{\mathsf{T}}\right) \\
& + 2\cdot\sum_{\iota_1=\mathsf{M}+1}^{\infty}\sum_{\iota_2=0}^{\infty} \operatorname{tr}\left(\big(\partial_i^{st}\mathsf{k}_{\iota_1}(\vartheta^0)\Delta Y_{k-\iota_1}^{(h)}\big)\big(V_{\vartheta^0}^{(h)}\big)^{-1}\Delta Y_{k-\iota_2}^{(h)\mathsf{T}}\mathsf{k}_{\iota_2}(\vartheta^0)^{\mathsf{T}}\right),
\end{aligned}$$

$$
\begin{aligned}
U_{\mathsf{M},k}^{(i)} := & \sum_{\iota_2=\mathsf{M}+1}^{\infty} \operatorname{tr}\Big(\big(V_{\vartheta^0}^{(h)}\big)^{-1} \Pi(\vartheta) Y_{st,k-1}^{(h)} \Delta Y_{k-\iota_2}^{(h)\mathsf{T}} \mathsf{k}_{\iota_2}(\vartheta^0)^{\mathsf{T}} \big(V_{\vartheta^0}^{(h)}\big)^{-1} \partial_i^{st} V_{\vartheta^0}^{(h)} \Big) \\
& - \sum_{\iota_1=0}^{\infty} \sum_{\iota_2=\mathsf{M}+1}^{\infty} \operatorname{tr}\Big(\big(V_{\vartheta^0}^{(h)}\big)^{-1} \mathsf{k}_{\iota_1}(\vartheta^0) \Delta Y_{k-\iota_1}^{(h)} \Delta Y_{k-\iota_2}^{(h)\mathsf{T}} \mathsf{k}_{\iota_2}(\vartheta^0)^{\mathsf{T}} \big(V_{\vartheta^0}^{(h)}\big)^{-1} \partial_i^{st} V_{\vartheta^0}^{(h)} \Big) \\
& - 2 \cdot \sum_{\iota_2=\mathsf{M}+1}^{\infty} \operatorname{tr}\Big(\big(\partial_i^{st} \Pi(\vartheta^0) Y_{st,k-1}^{(h)}\big) \big(V_{\vartheta^0}^{(h)}\big)^{-1} \Delta Y_{k-\iota_2}^{(h)\mathsf{T}} \mathsf{k}_{\iota_2}(\vartheta^0)^{\mathsf{T}} \Big) \\
& + 2 \cdot \sum_{\iota_1=0}^{\infty} \sum_{\iota_2=\mathsf{M}+1}^{\infty} \operatorname{tr}\Big(\big(\partial_i^{st} \mathsf{k}_{\iota_1}(\vartheta^0) \Delta Y_{k-\iota_1}^{(h)}\big) \big(V_{\vartheta^0}^{(h)}\big)^{-1} \Delta Y_{k-\iota_2}^{(h)\mathsf{T}} \mathsf{k}_{\iota_2}(\vartheta^0)^{\mathsf{T}} \Big).
\end{aligned}
$$

We define $\mathcal{Y}_{\mathsf{M},k} := (Y_{\mathsf{M},k}^{(1)\,\mathsf{T}}, \ldots, Y_{\mathsf{M},k}^{(s_2)\,\mathsf{T}})^{\mathsf{T}}$ as well as $\mathcal{Z}_{\mathsf{M},k} := (Z_{\mathsf{M},k}^{(1)\,\mathsf{T}}, \ldots, Z_{\mathsf{M},k}^{(s_2)\,\mathsf{T}})^{\mathsf{T}}$ and use a truncation argument analogous to Schlemm and Stelzer [51, Lemma 2.16]. The main difference to Schlemm and Stelzer [51] is that in our case the definition of $Y_{\mathsf{M},k}^{(i)}$, $V_{\mathsf{M},k}^{(i)}$ and $U_{\mathsf{M},k}^{(i)}$ are more complex including the two stochastic processes $\Delta Y^{(h)}$, $Y_{st}^{(h)}$ and additional summands. We show the claim in three steps.

Step 1: The process $\mathcal{Y}_{\mathsf{M},k}$ depends only on $\mathsf{M}+1$ past values of $\Delta Y^{(h)}$ and $Y_{st}^{(h)}$. Hence, it inherits the strong mixing property of $(\Delta Y^{(h)}, Y_{st}^{(h)})$ and satisfies

$$
\alpha_{\mathcal{Y}_{\mathsf{M}}}(l) \le \alpha_{\Delta Y^{(h)}, Y_{st}^{(h)}}(\max\{0, l-\mathsf{M}-1\}).
$$

Thus, by Lemma 4.5 we have $\sum_{l=1}^{\infty} \left(\alpha_{\mathcal{Y}_{\mathsf{M}}}(l)\right)^{\delta/(2+\delta)} < \infty$. Using the Cramér-Wold device and the univariate central limit theorem of Ibragimov [29, Theorem 1.7] for strongly mixing random variables we obtain

$$
\frac{1}{\sqrt{n}} \sum_{k=1}^{n} \left(\mathcal{Y}_{\mathsf{M},k} - \mathbb{E}\mathcal{Y}_{\mathsf{M},k}\right) \xrightarrow{w} \mathcal{N}(0_{s_2}, I_{\mathsf{M}}(\vartheta_2^0))
$$

as $n \to \infty$ where $I_{\mathsf{M}}(\vartheta_2^0) := \sum_{l \in \mathbb{Z}} \operatorname{Cov}(\mathcal{Y}_{\mathsf{M},1}, \mathcal{Y}_{\mathsf{M},1+l})$. Next, we need to show that

$$
I_{\mathsf{M}}(\vartheta_2^0) \xrightarrow{\mathsf{M}\to\infty} I(\vartheta_2^0). \tag{4.11}
$$

Therefore, we prove that $\operatorname{Cov}\left(Y_{\mathsf{M},k}^{(i)}, Y_{\mathsf{M},k+l}^{(j)}\right) \to \operatorname{Cov}\left(\partial_i^{st} \ell_{1,2}^{(h)}(\vartheta^0), \partial_j^{st} \ell_{1+l,2}^{(h)}(\vartheta^0)\right)$ as $\mathsf{M} \to \infty$. Note that the bilinearity property of the covariance operator implies

$$
\begin{aligned}
& |\operatorname{Cov}\left(Y_{\mathsf{M},k}^{(i)}, Y_{\mathsf{M},k+l}^{(j)}\right) - \operatorname{Cov}\left(\partial_i^{st} \ell_{k,2}^{(h)}(\vartheta^0), \partial_j^{st} \ell_{k+l,2}^{(h)}(\vartheta^0)\right)| \\
= \quad & |\operatorname{Cov}\left(Y_{\mathsf{M},k}^{(i)}, Y_{\mathsf{M},k+l}^{(j)} - \partial_j^{st} \ell_{k+l,2}^{(h)}(\vartheta^0)\right) \\
& + \operatorname{Cov}\left(Y_{\mathsf{M},k}^{(i)} - \partial_i^{st} \ell_{k,2}^{(h)}(\vartheta^0), \partial_j^{st} \ell_{k+l,2}^{(h)}(\vartheta^0)\right)| \\
\le \quad & \operatorname{Var}\left(Y_{\mathsf{M},1}^{(i)}\right)^{1/2} \operatorname{Var}\left(Y_{\mathsf{M},1}^{(j)} - \partial_j^{st} \ell_{1,2}^{(h)}(\vartheta^0)\right)^{1/2} \\
& + \operatorname{Var}\left(Y_{\mathsf{M},1}^{(i)} - \partial_i^{st} \ell_{1,2}^{(h)}(\vartheta^0)\right)^{1/2} \operatorname{Var}\left(\partial_j^{st} \ell_{l,2}^{(h)}(\vartheta^0)\right)^{1/2},
\end{aligned} \tag{4.12}
$$

where

$$
\begin{aligned}
&Y^{(i)}_{\mathsf{M},k} - \partial_i^{st}\ell^{(h)}_{k,2}(\vartheta^0)\\
&= -\sum_{\iota_1=\mathsf{M}+1}^{\infty} \operatorname{tr}\left(\left(V^{(h)}_{\vartheta^0}\right)^{-1}\mathsf{k}_{\iota_1}(\vartheta^0)\Delta Y^{(h)}_{k-\iota_1}Y^{(h)\mathsf{T}}_{st,k-1}\Pi(\vartheta^0)^{\mathsf{T}}\left(V^{(h)}_{\vartheta^0}\right)^{-1}\partial_i^{st}V^{(h)}_{\vartheta^0}\right)\\
&\quad -\sum_{\iota_2=\mathsf{M}+1}^{\infty} \operatorname{tr}\left(\left(V^{(h)}_{\vartheta^0}\right)^{-1}\Pi(\vartheta^0)Y^{(h)}_{st,k-1}\Delta Y^{(h)\mathsf{T}}_{k-\iota_2}\mathsf{k}_{\iota_2}(\vartheta^0)^{\mathsf{T}}\left(V^{(h)}_{\vartheta^0}\right)^{-1}\partial_i^{st}V^{(h)}_{\vartheta^0}\right)\\
&\quad +\sum_{\substack{\iota_1,\iota_2\\ \max\{\iota_1,\iota_2\}>\mathsf{M}}} \operatorname{tr}\left(\left(V^{(h)}_{\vartheta^0}\right)^{-1}\mathsf{k}_{\iota_1}(\vartheta^0)\Delta Y^{(h)}_{k-\iota_1}\Delta Y^{(h)\mathsf{T}}_{k-\iota_2}\mathsf{k}_{\iota_2}(\vartheta^0)^{\mathsf{T}}\left(V^{(h)}_{\vartheta^0}\right)^{-1}\partial_i^{st}V^{(h)}_{\vartheta^0}\right)\\
&\quad +2\cdot\sum_{\iota_1=\mathsf{M}+1}^{\infty} \operatorname{tr}\left(\left(\partial_i^{st}\mathsf{k}_{\iota_1}(\vartheta^0)\Delta Y^{(h)}_{k-\iota_1}\right)\left(V^{(h)}_{\vartheta^0}\right)^{-1}Y^{(h)\mathsf{T}}_{st,k-1}\Pi(\vartheta^0)^{\mathsf{T}}\right)\\
&\quad +2\cdot\sum_{\iota_2=\mathsf{M}+1}^{\infty} \operatorname{tr}\left(\left(\partial_i^{st}\Pi(\vartheta^0)Y^{(h)}_{st,k-1}\right)\left(V^{(h)}_{\vartheta^0}\right)^{-1}\Delta Y^{(h)\mathsf{T}}_{k-\iota_2}\mathsf{k}_{\iota_2}(\vartheta^0)^{\mathsf{T}}\right)\\
&\quad -2\cdot\sum_{\substack{\iota_1,\iota_2\\ \max\{\iota_1,\iota_2\}>\mathsf{M}}} \operatorname{tr}\left(\left(\partial_i^{st}\mathsf{k}_{\iota_1}(\vartheta^0)\Delta Y^{(h)}_{k-\iota_1}\right)\left(V^{(h)}_{\vartheta^0}\right)^{-1}\Delta Y^{(h)\mathsf{T}}_{k-\iota_2}\mathsf{k}_{\iota_2}(\vartheta^0)^{\mathsf{T}}\right).
\end{aligned}
$$

We obtain with the Cauchy-Schwarz inequality, the exponentially decreasing coefficients $(\mathsf{k}_j(\vartheta^0))_{j\in\mathbb{N}}$ and the finite 4th-moment of $Y^{(h)}_{st}$ and $\Delta Y^{(h)}$ due to Assumption A that for some $0<\rho<1$,

$$\operatorname{Var}\left(Y^{(i)}_{\mathsf{M},1} - \partial_i^{st}\ell^{(h)}_{1,2}(\vartheta^0)\right) \le \mathfrak{C}\rho^{\mathsf{M}}.$$

Moreover, by the proof of Lemma 4.6 we have $\operatorname{Var}\left(\partial_j^{st}\ell^{(h)}_{1,2}(\vartheta^0)\right)<\infty$ and then, $\operatorname{Var}\left(Y^{(i)}_{\mathsf{M},1}\right) \le 2\mathbb{E}\left(Y^{(i)}_{\mathsf{M},1} - \partial_i^{st}\ell^{(h)}_{1,2}(\vartheta^0)\right)^2 + 2\mathbb{E}\left(\partial_j^{st}\ell^{(h)}_{1,2}(\vartheta^0)\right)^2 < \infty$ as well. Thus, (4.12) converges uniformly in l at an exponential rate to zero as $\mathsf{M}\to\infty$ and

$$\operatorname{Cov}\left(Y^{(i)}_{\mathsf{M},k}, Y^{(j)}_{\mathsf{M},k+l}\right) \xrightarrow{\mathsf{M}\to\infty} \operatorname{Cov}\left(\partial_i^{st}\ell^{(h)}_{1,2}(\vartheta^0), \partial_j^{st}\ell^{(h)}_{1+l,2}(\vartheta^0)\right).$$

Then, the same arguments as in Schlemm and Stelzer [51, Lemma 2.16] guarantee that there exists a dominant (see Scholz [52, Section 5.9]) so that dominated convergence results in (4.11) (see proof of Lemma 4.7).

Step 2: In this step, we show that $\frac{1}{\sqrt{n}}\sum_{k=1}^n\left(\mathcal{Z}_{\mathsf{M},k} - \mathbb{E}\mathcal{Z}_{\mathsf{M},k}\right)$ is asymptotically negligible. We have

$$
\begin{aligned}
&\operatorname{tr}\left(\operatorname{Var}\left(\frac{1}{\sqrt{n}}\sum_{k=1}^n \mathcal{Z}_{\mathsf{M},k}\right)\right)\\
&\qquad \le 2\cdot\operatorname{tr}\left(\operatorname{Var}\left(\frac{1}{\sqrt{n}}\sum_{k=1}^n \mathcal{U}_{\mathsf{M},k}\right)\right) + 2\cdot\operatorname{tr}\left(\operatorname{Var}\left(\frac{1}{\sqrt{n}}\sum_{k=1}^n \mathcal{V}_{\mathsf{M},k}\right)\right), \qquad (4.13)
\end{aligned}
$$

where $\mathcal{U}_{\mathsf{M},k}$ and $\mathcal{V}_{\mathsf{M},k}$ are defined in the same vein as $\mathcal{Z}_{\mathsf{M},k}$. Since both terms can be treated similarly we consider only the first one

$$\begin{aligned}\operatorname{tr}\left(\operatorname{Var}\left(\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\mathcal{U}_{\mathsf{M},k}\right)\right) &= \frac{1}{n}\sum_{k,k'=1}^{n}\operatorname{tr}\left(\operatorname{Cov}(\mathcal{U}_{\mathsf{M},k},\mathcal{U}_{\mathsf{M},k'})\right)\\ &\leq \sum_{i,j=1}^{s_2}\sum_{l=-\infty}^{\infty}|\operatorname{Cov}(U_{\mathsf{M},1}^{(i)},U_{\mathsf{M},1+l}^{(j)})|. \end{aligned} \tag{4.14}$$

With the same arguments as in Schlemm and Stelzer [51, Lemma 2.16] we obtain independent of i and j the upper bound

$$\begin{aligned}\sum_{l=0}^{\infty}|\operatorname{Cov}(U_{\mathsf{M},1}^{(i)},U_{\mathsf{M},1+l}^{(j)})| &\leq \sum_{l=0}^{2\mathsf{M}}|\operatorname{Cov}(U_{\mathsf{M},1}^{(i)},U_{\mathsf{M},1+l}^{(j)})|+\sum_{l=2\mathsf{M}+1}^{\infty}|\operatorname{Cov}(U_{\mathsf{M},1}^{(i)},U_{\mathsf{M},1+l}^{(j)})|\\ &\leq \mathfrak{C}\rho^{\mathsf{M}}\left(\mathsf{M}+\sum_{l=0}^{\infty}\left[\alpha_{\Delta Y^{(h)},Y_{st}^{(h)}}(l)\right]^{\frac{\delta}{\delta+2}}\right),\end{aligned}$$

which implies $\operatorname{tr}\left(\operatorname{Var}\left(\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\mathcal{U}_{\mathsf{M},k}\right)\right)\leq \mathfrak{C}\rho^{\mathsf{M}}(\mathsf{M}+\overline{\mathfrak{C}})$, due to (4.14), for some constant $\overline{\mathfrak{C}}>0$. With the same ideas one obtains an equivalent bound for $\operatorname{tr}\left(\operatorname{Var}\left(\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\mathcal{V}_{\mathsf{M},k}\right)\right)$ and thus, we have with (4.13) that

$$\operatorname{tr}\left(\operatorname{Var}\left(\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\mathcal{Z}_{\mathsf{M},k}\right)\right)\leq \mathfrak{C}\rho^{\mathsf{M}}(\mathsf{M}+\overline{\mathfrak{C}}). \tag{4.15}$$

Step 3: With the multivariate Chebyshev inequality (see Schlemm [49, Lemma 3.19]) and (4.15) from Step 2 we obtain for every $\tau>0$ that

$$\begin{aligned}&\lim_{\mathsf{M}\to\infty}\limsup_{n\to\infty}\mathbb{P}\left(\left\|\sqrt{n}\nabla_{\vartheta_2}\mathcal{L}_n^{(h)}(\vartheta^0)-\frac{1}{\sqrt{n}}\sum_{k=1}^{n}[\mathcal{Y}_{\mathsf{M},k}-\mathbb{E}\mathcal{Y}_{\mathsf{M},k}]\right\|>\tau\right)\\ &\quad\leq \lim_{\mathsf{M}\to\infty}\limsup_{n\to\infty}\frac{s_2}{\tau^2}\operatorname{tr}\left(\operatorname{Var}\left(\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\mathcal{Z}_{\mathsf{M},k}\right)\right)\leq \lim_{\mathsf{M}\to\infty}\frac{s_2}{\tau^2}\mathfrak{C}\rho^{\mathsf{M}}(\mathsf{M}+\overline{\mathfrak{C}})=0.\end{aligned}$$

All in all, the results of Step 1 and Step 3 combined with Brockwell and Davis [13, Proposition 6.3.9] yield the asymptotic normality in Lemma 4.8. □

4.2.2. Asymptotic behavior of the Hessian matrix

We require an additional assumption for the Hessian matrix (with respect to the short-run parameters) to be positive definite. Therefore, we need some notation. We write shortly $F_\vartheta := \mathrm{e}^{A_\vartheta h}-K_\vartheta^{(h)}C_\vartheta$. The function is similar to the function in Schlemm and Stelzer [51, Assumption C11]. However, we define F_ϑ different since we do not have a moving average representation of $Y^{(h)}$ with respect to the innovations $\varepsilon^{(h)}$. Hence, we have to adapt the criterion in Schlemm and Stelzer [51] and define the function

$$\psi_{\vartheta,j} := \left(\begin{matrix} \left[I_{j+1} \otimes K_\vartheta^{(h)\,\mathsf{T}} \otimes C_\vartheta\right] \left[(\mathrm{vec}\ I_N)^\mathsf{T} \quad (\mathrm{vec}\ F_\vartheta)^\mathsf{T} \ \ldots \ (\mathrm{vec}\ F_\vartheta^j)^\mathsf{T}\right]^\mathsf{T} \\ \mathrm{vec}\ V_\vartheta^{(h)} \end{matrix} \right). \tag{4.16}$$

Assumption F. *Let there exist a positive index j_0 such that the $[(j_0+2)d^2 \times s_2]$ matrix $\nabla_{\vartheta_2}\psi_{\vartheta^0,j_0}$ has rank s_2.*

Proposition 4.9. *Let Assumption F additionally hold. Then,*

$$\nabla^2_{\vartheta_2}\mathcal{L}_n^{(h)}(\vartheta^0) \xrightarrow{p} Z_{st}(\vartheta^0),$$

where the $(s_2 \times s_2)$-dimensional matrix $Z_{st}(\vartheta^0)$ is given by

$$\begin{aligned} [Z_{st}(\vartheta^0)]_{i,j} =& 2\mathbb{E}\big(\partial_i^{st}\varepsilon_1^{(h)}(\vartheta^0)^\mathsf{T}\big)\big(V_{\vartheta^0}^{(h)}\big)^{-1}\big(\partial_j^{st}\varepsilon_1^{(h)}(\vartheta^0)\big) \\ &+ \mathrm{tr}\left(\big(V_{\vartheta^0}^{(h)}\big)^{-\frac{1}{2}}\big(\partial_i^{st}V_{\vartheta^0}^{(h)}\big)\big(V_{\vartheta^0}^{(h)}\big)^{-1}\partial_j^{st}V_{\vartheta^0}^{(h)}\big(V_{\vartheta^0}^{(h)}\big)^{-\frac{1}{2}}\right). \end{aligned}$$

Moreover, the limiting matrix $Z_{st}(\vartheta^0)$ is almost surely a non-singular deterministic matrix.

Proof. We proceed as in the proof of Proposition 4.2.
Step 1: Since $(\partial_{i,j}^{st}\varepsilon_k^{(h)}(\vartheta^0), \partial_j^{st}\varepsilon_k^{(h)}(\vartheta^0), \varepsilon_k^{(h)}(\vartheta^0))_{k\in\mathbb{N}}$ is a stationary and an ergodic sequence with finite absolute moment (see Lemma B.3(a)) we obtain with Birkhoff's Ergodic Theorem

$$\begin{aligned} &\partial_{i,j}^{st}\mathcal{L}_n^{(h)}(\vartheta^0) \\ &\xrightarrow{p} \mathrm{tr}\left(\big(V_{\vartheta^0}^{(h)}\big)^{-1}\partial_{i,j}^{st}V_{\vartheta^0}^{(h)} - \big(V_{\vartheta^0}^{(h)}\big)^{-1}\big(\partial_i^{st}V_{\vartheta^0}^{(h)}\big)\big(V_{\vartheta^0}^{(h)}\big)^{-1}\big(\partial_j^{st}V_{\vartheta^0}^{(h)}\big)\right) \\ &\quad - \mathrm{tr}\left(\big(V_{\vartheta^0}^{(h)}\big)^{-1}\mathbb{E}\left[\varepsilon_1^{(h)}(\vartheta^0)\varepsilon_1^{(h)}(\vartheta^0)^\mathsf{T}\right]\big(V_{\vartheta^0}^{(h)}\big)^{-1}\partial_{i,j}^{st}V_{\vartheta^0}^{(h)}\right) \\ &\quad + \mathrm{tr}\left(\big(V_{\vartheta^0}^{(h)}\big)^{-1}\big(\partial_j^{st}V_{\vartheta^0}^{(h)}\big)\big(V_{\vartheta^0}^{(h)}\big)^{-1}\mathbb{E}\left[\varepsilon_1^{(h)}(\vartheta^0)\varepsilon_1^{(h)}(\vartheta^0)^\mathsf{T}\right]\big(V_{\vartheta^0}^{(h)}\big)^{-1}\partial_i^{st}V_{\vartheta^0}^{(h)}\right) \\ &\quad + \mathrm{tr}\left(\big(V_{\vartheta^0}^{(h)}\big)^{-1}\mathbb{E}\left[\varepsilon_1^{(h)}(\vartheta^0)\varepsilon_1^{(h)}(\vartheta^0)^\mathsf{T}\right]\big(V_{\vartheta^0}^{(h)}\big)^{-1}\big(\partial_j^{st}V_{\vartheta^0}^{(h)}\big)\big(V_{\vartheta^0}^{(h)}\big)^{-1}\partial_i^{st}V_{\vartheta^0}^{(h)}\right) \\ &\quad - \mathrm{tr}\left(\big(V_{\vartheta^0}^{(h)}\big)^{-1}\mathbb{E}\left[\partial_j^{st}\varepsilon_1^{(h)}(\vartheta^0)\varepsilon_1^{(h)}(\vartheta^0)^\mathsf{T}\right]\big(V_{\vartheta^0}^{(h)}\big)^{-1}\partial_i^{st}V_{\vartheta^0}^{(h)}\right) \\ &\quad +2\cdot\mathrm{tr}\left(\big(V_{\vartheta^0}^{(h)}\big)^{-1}\mathbb{E}\left[\varepsilon_1^{(h)}(\vartheta^0)\left(\partial_{i,j}^{st}\varepsilon_1^{(h)}(\vartheta^0)^\mathsf{T}\right)\right]\right) \\ &\quad -2\cdot\mathrm{tr}\left(\big(V_{\vartheta^0}^{(h)}\big)^{-1}\mathbb{E}\left[\varepsilon_1^{(h)}(\vartheta^0)\partial_i^{st}\varepsilon_1^{(h)}(\vartheta^0)^\mathsf{T}\right]\big(V_{\vartheta^0}^{(h)}\big)^{-1}\partial_j^{st}V_{\vartheta^0}^{(h)}\right) \\ &\quad +2\cdot\mathbb{E}\left[\left(\partial_i^{st}\varepsilon_1^{(h)}(\vartheta^0)^\mathsf{T}\right)\big(V_{\vartheta^0}^{(h)}\big)^{-1}\left(\partial_j^{st}\varepsilon_1^{(h)}(\vartheta^0)\right)\right]. \end{aligned}$$

Combining this with Lemma B.3(c,d) results in

$$\partial_{i,j}^{st}\mathcal{L}_n^{(h)}(\vartheta^0) \quad \xrightarrow{p} \quad \mathrm{tr}\left(\big(V_{\vartheta^0}^{(h)}\big)^{-1}\big(\partial_i^{st}V_{\vartheta^0}^{(h)}\big)\big(V_{\vartheta^0}^{(h)}\big)^{-1}\partial_j^{st}V_{\vartheta^0}^{(h)}\right)$$

$$+2\cdot\mathbb{E}\left[\big(\partial_i^{st}\varepsilon_1^{(h)}(\vartheta^0)^\mathsf{T}\big)\big(V_{\vartheta^0}^{(h)}\big)^{-1}\big(\partial_j^{st}\varepsilon_1^{(h)}(\vartheta^0)\big)\right].$$

Step 2: Next we check that $Z_{st}(\vartheta^0)$ is positive definite with probability one. That we show by contradiction similarly to the corresponding proofs in Schlemm and Stelzer [51, Lemma 3.22] or Boubacar and Francq [11, Lemma 4], respectively. From Step 2 we know that

$$\nabla^2_{\vartheta_2}\mathcal{L}_n^{(h)}(\vartheta^0)\xrightarrow{p} Z_{st}(\vartheta^0)=Z_{st,1}(\vartheta^0)+Z_{st,2}(\vartheta^0), \tag{4.17}$$

where $Z_{st,1}(\vartheta^0):=2\cdot\left[\mathbb{E}\left(\partial_i^{st}\varepsilon_1^{(h)}(\vartheta^0)^\mathsf{T}\right)\left(V_{\vartheta^0}^{(h)}\right)^{-1}\left(\partial_j^{st}\varepsilon_1^{(h)}(\vartheta^0)\right)\right]_{1\le i,j\le s_2}$

and $Z_{st,2}(\vartheta^0):=\left[\operatorname{tr}\left(\big(V_{\vartheta^0}^{(h)}\big)^{-\frac12}\big(\partial_i^{st}V_{\vartheta^0}^{(h)}\big)\big(V_{\vartheta^0}^{(h)}\big)^{-1}\partial_j^{st}V_{\vartheta^0}^{(h)}\big(V_{\vartheta^0}^{(h)}\big)^{-\frac12}\right)\right]_{1\le i,j\le s_2}.$

We can factorize $Z_{st,2}(\vartheta^0)$ in the following way:

$$Z_{st,2}(\vartheta^0)=\begin{pmatrix}a_1 & \dots & a_{s_2}\end{pmatrix}^\mathsf{T}\begin{pmatrix}a_1 & \dots & a_{s_2}\end{pmatrix}\qquad\text{with}$$
$$a_i:=\left(\big(V_{\vartheta^0}^{(h)}\big)^{-\frac12}\otimes\big(V_{\vartheta^0}^{(h)}\big)^{-\frac12}\right)\operatorname{vec}\big(\partial_i^{st}V_{\vartheta^0}^{(h)}\big).$$

Thus, $Z_{st,2}(\vartheta^0)$ is obviously positive semi-definite. Similarly, $Z_{st,1}(\vartheta^0)$ is positive semi-definite. It remains to check that for any $b\in\mathbb{R}^{s_2}\setminus\{0_{s_2}\}$ we have $b^\mathsf{T}Z_{st,1}(\vartheta^0)b+b^\mathsf{T}Z_{st,2}(\vartheta^0)b>0$. We assume for the sake of contradiction that there exists a vector $b\in\mathbb{R}^{s_2}\setminus\{0_{s_2}\}$ such that

$$b^\mathsf{T}Z_{st,1}(\vartheta^0)b+b^\mathsf{T}Z_{st,2}(\vartheta^0)b=0. \tag{$\Diamond$}$$

In order to be zero, each summand $b^\mathsf{T}Z_{st,1}(\vartheta^0)b$ and $b^\mathsf{T}Z_{st,2}(\vartheta^0)b$ must be zero, since $Z_{st,1}(\vartheta^0)$ as well as $Z_{st,2}(\vartheta^0)$ are positive semi-definite. But $b^\mathsf{T}Z_{st,1}(\vartheta^0)b=0$ is only possible if

$$0_d=(\nabla_{\vartheta_2}\varepsilon_1^{(h)}(\vartheta^0))b=-\sum_{j=1}^{\infty}\left(\nabla_{\vartheta_2}\left[C_{\vartheta^0}F_{\vartheta^0}^{j-1}K_{\vartheta^0}^{(h)}Y_{k-j}^{(h)}\right]\right)b\quad\mathbb{P}\text{-a.s.}$$

Rewriting this equation yields

$$\left(\nabla_{\vartheta_2}\left[C_{\vartheta^0}K_{\vartheta^0}^{(h)}Y_{k-1}^{(h)}\right]\right)b=-\sum_{j=2}^{\infty}\left(\nabla_{\vartheta_2}\left[C_{\vartheta^0}F_{\vartheta^0}^{j-1}K_{\vartheta^0}^{(h)}Y_{k-j}^{(h)}\right]\right)b\quad\mathbb{P}\text{-a.s.} \tag{4.18}$$

Hence, for every row $i=1,\dots,d$ and $b=(b_1,\dots,b_{s_2})^\mathsf{T}$ we obtain

$$\sum_{u=1}^{s_2}\left[\sum_{l=1}^{d}\partial_u^{st}(C_{\vartheta^0}K_{\vartheta^0}^{(h)})_{i,l}Y_{k-1,l}^{(h)}\right]b_u$$
$$=-\sum_{j=2}^{\infty}\sum_{u=1}^{s_2}\left[\sum_{l=1}^{d}\partial_u^{st}(C_{\vartheta^0}F_{\vartheta^0}^{j-1}K_{\vartheta^0}^{(h)})_{i,l}Y_{k-j,l}^{(h)}\right]b_u\quad\mathbb{P}\text{-a.s.},$$

which is equivalent to

$$\left(\nabla_{\vartheta_2}\left[e_i^\top C_{\vartheta^0} K_{\vartheta^0}^{(h)}\right] b\right)^\top Y_{k-1}^{(h)} = -\sum_{j=2}^{\infty}\left(\nabla_{\vartheta_2}\left[e_i^\top C_{\vartheta^0} F_{\vartheta^0}^{j-1} K_{\vartheta^0}^{(h)}\right] b\right)^\top Y_{k-j}^{(h)} \quad \mathbb{P}\text{-a.s.}$$

But then $\left(\nabla_{\vartheta_2}\left[e_i^\top C_{\vartheta^0} K_{\vartheta^0}^{(h)}\right] b\right)^\top Y_{k-1}^{(h)}$ lies in $\overline{\text{span}}\{Y_j^{(h)} : j \leq k-2\}$. By the definition of the linear innovations, this is only possible if $\left(\nabla_{\vartheta_2}\left[e_i^\top C_{\vartheta^0} K_{\vartheta^0}^{(h)}\right] b\right)^\top \varepsilon_{k-1}^{(h)} = 0$ $\mathbb{P}$-a.s. However, $V_{\vartheta^0}^{(h)} = \mathbb{E}(\varepsilon_{k-1}^{(h)}(\varepsilon_{k-1}^{(h)})^\top)$ is non-singular due to Scholz [52, Lemma 5.9.1] so that necessarily $\nabla_{\vartheta_2}\left[e_i^\top C_{\vartheta^0} K_{\vartheta^0}^{(h)}\right] b = 0_d$ for $i = 1, \ldots, d$. This is again equivalent to $\nabla_{\vartheta_2}(C_{\vartheta^0} K_{\vartheta^0}^{(h)})b = 0_{d^2}$. Plugging this in (4.18) gives

$$\nabla_{\vartheta_2}\left[C_{\vartheta^0} F_{\vartheta^0} K_{\vartheta^0}^{(h)} Y_{k-2}^{(h)}\right] b = -\sum_{j=3}^{\infty} \nabla_{\vartheta_2}\left[C_{\vartheta^0} F_{\vartheta^0}^{j-1} K_{\vartheta^0}^{(h)} Y_{k-j}^{(h)}\right] b.$$

Then, we can show similarly $\nabla_{\vartheta_2}(C_{\vartheta^0} F_{\vartheta^0} K_{\vartheta^0}^{(h)})b = 0_{d^2}$ and obtain recursively that

$$\nabla_{\vartheta_2}(C_{\vartheta^0} F_{\vartheta^0}^j K_{\vartheta^0}^{(h)})b = 0_{d^2} \quad \text{for } j \in \mathbb{N}_0. \tag{4.19}$$

On the other hand, we obtain due $b^\top Z_{st,2}(\vartheta^0)b = 0$ under assumption ($\Diamond$) that

$$\left(\nabla_{\vartheta_2} V_{\vartheta^0}^{(h)}\right)b = 0_{d^2}. \tag{4.20}$$

The definition of $\psi_{\vartheta,j}$ in (4.16), (4.19) and (4.20) imply that $\left(\nabla_{\vartheta_2}\psi_{\vartheta^0,j}\right)b = 0_{(j+2)d^2}$ holds for all $j \in \mathbb{N}$, which contradicts Assumption F. Hence, $Z_{st}(\vartheta^0)$ is almost surely positive definite. □

4.2.3. *Asymptotic normality of the short-run QML estimator*

We conclude this section with the last main result of this paper, namely the asymptotic distribution of the short-run QML estimator.

Theorem 4.10. *Let Assumption F additionally hold. Furthermore, suppose*

$$I(\vartheta^0) = \lim_{n\to\infty} \text{Var}\left(\nabla_{\vartheta_2}\mathcal{L}_n^{(h)}(\vartheta^0)\right) \quad \textit{and} \quad Z_{st}(\vartheta^0) = \lim_{n\to\infty} \nabla_{\vartheta_2}^2 \mathcal{L}_n^{(h)}(\vartheta^0).$$

Then, as $n \to \infty$,

$$\sqrt{n}(\widehat{\vartheta}_{n,2} - \vartheta_2^0) \xrightarrow{w} \mathcal{N}(0, Z_{st}(\vartheta^0)^{-1} I(\vartheta^0) Z_{st}(\vartheta^0)^{-1}).$$

Again we need the following auxiliary result for the proof.

Lemma 4.11. *For every $\tau > 0$ and every $\eta > 0$, there exist an integer $n(\tau,\eta)$ and real numbers $\delta_1, \delta_2 > 0$ such that for $\frac{3}{4} < \gamma < 1$ and $n \geq n(\tau,\eta)$:*

(a) $\mathbb{P}\left(\sup_{\vartheta_1 \in N_{n,\gamma}(\vartheta_1^0,\delta_1)} \|\sqrt{n}\nabla_{\vartheta_2}\mathcal{L}_n^{(h)}(\vartheta_1,\vartheta_2^0) - \sqrt{n}\nabla_{\vartheta_2}\mathcal{L}_n^{(h)}(\vartheta_1^0,\vartheta_2^0)\| > \tau \right) \leq \eta,$

(b) $\mathbb{P}\left(\sup_{\vartheta \in N_{n,\gamma}(\vartheta_1^0,\delta_1)\times \mathcal{B}(\vartheta_2^0,\delta_2)} \|\nabla^2_{\vartheta_2}\mathcal{L}_n^{(h)}(\vartheta) - \nabla^2_{\vartheta_2}\mathcal{L}_n^{(h)}(\vartheta^0)\| > \tau \right) \leq \eta.$

This local stochastic equicontinuity condition for the standardized score $\sqrt{n}\nabla_{\vartheta_2}\mathcal{L}_n^{(h)}(\cdot,\vartheta_2^0)$ in ϑ_1^0 and for the standardized Hessian matrix $\nabla^2_{\vartheta_2}\mathcal{L}_n^{(h)}(\cdot)$ in ϑ^0 do not hold for general ϑ_1 and ϑ, respectively. Accordingly the stochastic equicontinuity conditions of Saikkonen [47] are not satisfied.

Proof of Lemma 4.11.
(a) We use the upper bound

$$\begin{aligned}
&\sup_{\vartheta_1 \in N_{n,\gamma}(\vartheta_1^0,\delta_1)} \|\sqrt{n}\nabla_{\vartheta_2}\mathcal{L}_n^{(h)}(\vartheta_1,\vartheta_2^0) - \sqrt{n}\nabla_{\vartheta_2}\mathcal{L}_n^{(h)}(\vartheta_1^0,\vartheta_2^0)\| \\
&\quad \leq \sup_{\vartheta_1 \in N_{n,\gamma}(\vartheta_1^0,\delta_1)} \|\sqrt{n}\nabla_{\vartheta_2}\mathcal{L}_{n,1,1}^{(h)}(\vartheta_1,\vartheta_2^0) - \sqrt{n}\nabla_{\vartheta_2}\mathcal{L}_{n,1,1}^{(h)}(\vartheta_1^0,\vartheta_2^0)\| \\
&\qquad + \sup_{\vartheta_1 \in N_{n,\gamma}(\vartheta_1^0,\delta_1)} \|\sqrt{n}\nabla_{\vartheta_2}\mathcal{L}_{n,1,2}^{(h)}(\vartheta_1,\vartheta_2^0) - \sqrt{n}\nabla_{\vartheta_2}\mathcal{L}_{n,1,2}^{(h)}(\vartheta_1^0,\vartheta_2^0)\|.
\end{aligned} \tag{4.21}$$

Since $\Pi(\vartheta^0)C_1 = 0_{d\times c}$ and $\nabla_{\vartheta_2}(\Pi(\vartheta^0)C_1) = 0_{dc\times s_2}$ we can apply (A.3) and receive

$$\begin{aligned}
&\sup_{\vartheta_1 \in N_{n,\gamma}(\vartheta_1^0,\delta_1)} \|\sqrt{n}\nabla_{\vartheta_2}\mathcal{L}_{n,1,1}^{(h)}(\vartheta_1,\vartheta_2^0) - \sqrt{n}\nabla_{\vartheta_2}\mathcal{L}_{n,1,1}^{(h)}(\vartheta_1^0,\vartheta_2^0)\| \\
&\quad \leq \mathfrak{C} \sup_{\vartheta_1 \in N_{n,\gamma}(\vartheta_1^0,\delta_1)} n^{\frac{3}{2}}\|\vartheta_1 - \vartheta_1^0\|^2 \operatorname{tr}\left(\frac{1}{n^2}\sum_{k=1}^n L_{k-1}^{(h)}[L_{k-1}^{(h)}]^\mathsf{T}\right) \\
&\quad \leq \mathfrak{C} n^{\frac{3}{2}-2\gamma} \operatorname{tr}\left(\frac{1}{n^2}\sum_{k=1}^n L_{k-1}^{(h)}[L_{k-1}^{(h)}]^\mathsf{T}\right) \xrightarrow{p} 0,
\end{aligned} \tag{4.22}$$

where we used $\gamma > 3/4$ and $\operatorname{tr}\left(\frac{1}{n^2}\sum_{k=1}^n L_{k-1}^{(h)}[L_{k-1}^{(h)}]^\mathsf{T}\right) = \mathcal{O}_p(1)$ due to Proposition A.1(b). For the second term we get by (A.4) and (A.6), and similar calculations as in Lemma 3.3 that

$$\begin{aligned}
&\sup_{\vartheta_1 \in N_{n,\gamma}(\vartheta_1^0,\delta_1)} \|\sqrt{n}\nabla_{\vartheta_2}\mathcal{L}_{n,1,2}^{(h)}(\vartheta_1,\vartheta_2^0) - \sqrt{n}\nabla_{\vartheta_2}\mathcal{L}_{n,1,2}^{(h)}(\vartheta_1^0,\vartheta_2^0)\| \\
&\qquad \leq \sup_{\vartheta_1 \in N_{n,\gamma}(\vartheta_1^0,\delta_1)} \sqrt{n}\mathfrak{C}\|\vartheta_1 - \vartheta_1^0\| U_n \leq \mathfrak{C} n^{\frac{1}{2}-\gamma} U_n \xrightarrow{p} 0
\end{aligned} \tag{4.23}$$

due to $\gamma > 3/4$ and $U_n = \mathcal{O}_p(1)$. A combination of (4.21)-(4.23) proves (a).
(b) The proof is similar to (a). □

Proof of Theorem 4.10. The proof is similar to the proof of Theorem 4.3 using Proposition 4.8, Proposition 4.9, Proposition 2.5 and Lemma 4.11. □

5. Simulation study

In this section we want to demonstrate the validity of the proposed QML-method by a simulation study. The simulated state space processes are driven either by a standard Brownian motion or by a NIG (normal inverse Gaussian) Lévy process with mean value 0_m. The increment of an m-dimensional NIG Lévy process $L(t) - L(t-1)$ has the density

$$f_{NIG}(x;\mu,\alpha,\beta,\delta,\Delta) = \frac{\delta \mathrm{e}^{\delta\kappa}}{2\pi} \cdot \frac{\mathrm{e}^{\langle \beta, x\rangle}(1+\alpha g(x))}{\mathrm{e}^{\alpha g(x)} g(x)^3}, \quad x \in \mathbb{R}^m,$$
$$\text{where} \quad g(x) = \sqrt{\delta^2 + \langle x-\mu, \Delta(x-\mu)\rangle} \quad \text{and} \quad \kappa^2 = \alpha^2 - \langle \beta, \Delta\beta\rangle > 0,$$

$\mu \in \mathbb{R}^m$ is a location parameter, $\alpha \geq 0$ is a shape parameter, $\beta \in \mathbb{R}^m$ is a symmetry parameter, $\delta \geq 0$ is a scale parameter and $\Delta \in \mathbb{R}^{m\times m}$ is positive semi-definite with $\det \Delta = 1$ determining the dependence of the components of $(L(t))_{t\geq 0}$. The covariance of the process is then

$$\Sigma_L = \delta(\alpha - \beta^{\mathsf{T}}\Delta\beta)^{-\frac{1}{2}} \left(\Delta + (\alpha^2 - \beta^{\mathsf{T}}\Delta\beta)^{-1}\Delta\beta\beta^T\Delta\right).$$

For more details on NIG Lévy processes see, e.g., Barndorff-Nielsen [3]. In all simulation studies we have simulated 350 independent replications of a cointegrated state space process on an equidistant time grid $0, 0.01, \ldots, 2000$ by applying an Euler scheme to the stochastic differential equation (1.1) with initial value $X(0) = 0_N$ and h in the observation scheme is chosen as 1.

Moreover, we use canonical representations of the state space models. On the one hand, C_{1,ϑ_1} are chosen on such a way that C_{1,ϑ_1} are lower triangular matrices with $C_{1,\vartheta_1}^{\mathsf{T}} C_{1,\vartheta_1} = I_c$ and similarly $C_{1,\vartheta_1}^{\perp}$ are lower triangular matrices with $C_{1,\vartheta_1}^{\perp\mathsf{T}} C_{1,\vartheta_1}^{\perp} = I_{d-c}$ satisfying Assumption A, Assumption C, and Assumption E for a properly chosen parameter space Θ. On the other hand, the parametrization of the stationary part $Y_{st,\vartheta}$ is based on the echelon canonical form as given in Schlemm and Stelzer [51] such that as well Assumption A and Assumption D are satisfied for the properly chosen parameter space Θ. The echelon canonical form is widely used in the VARMA context, see, e.g., Lütkepohl and Poskitt [37] and the textbooks of Lütkepohl [35], or Hannan and Deistler [24]. In the context of linear state space models canonical representations can also be found in Guidorzi [23]. For the asymptotic normality of the short-run parameters we require additionally Assumption F. However, this condition cannot be checked analytically, this is only possible numerically.

5.1. *Bivariate state space model*

As canonical parametrization of the family of cointegrated state space models we take

$$A_{2,\vartheta} = \begin{pmatrix} \vartheta_1 & \vartheta_2 & 0 \\ 0 & 0 & 1 \\ \vartheta_3 & \vartheta_4 & \vartheta_5 \end{pmatrix}, \; B_{2,\vartheta} = \begin{pmatrix} \vartheta_1 & \vartheta_2 \\ \vartheta_6 & \vartheta_7 \\ \vartheta_3 + \vartheta_5\vartheta_6 & \vartheta_4 + \vartheta_5\vartheta_7 \end{pmatrix}, \; B_{1,\vartheta} = \begin{pmatrix} \vartheta_8 & \vartheta_9 \end{pmatrix},$$

$$\operatorname{vech}(\Sigma^L_\vartheta) = (\vartheta_{10}, \vartheta_{11}, \vartheta_{12}), \quad C_{1,\vartheta} = \begin{pmatrix} \frac{\vartheta_{13}^2 - 1}{\vartheta_{13}^2 + 1} \\ \frac{2 \cdot \vartheta_{13}}{\vartheta_{13}^2 + 1} \end{pmatrix}, \quad C_{2,\vartheta} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

This implies that we have one cointegration relation and the cointegration rank is equal to 1. In total we have 13 parameters. We use

$$\vartheta^0 = \begin{pmatrix} -1 & -2 & 1 & -2 & -3 & 1 & 2 & 1 & 1 & 0.4751 & -0.1622 & 0.3708 & 3 \end{pmatrix}^\mathsf{T}.$$

In order to obtain the covariance matrix of the NIG Lévy process, we have to set the parameters of the NIG Lévy process to

$$\delta = 1, \quad \alpha = 3, \quad \beta = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \Delta = \begin{pmatrix} 1.2 & -0.5 \\ -0.5 & 1 \end{pmatrix} \quad \text{and} \quad \mu = -\frac{1}{2\sqrt{31}} \begin{pmatrix} 3 \\ 2 \end{pmatrix}.$$

On this way the parameters of the stationary process $Y_{st,\vartheta}$ are chosen as in Schlemm and Stelzer [51, Section 4.2], who performed a simulation study for QML estimation of stationary state space processes.

TABLE 1
Sample mean, bias and sample standard deviation of 350 replications of the QML estimator for a two-dimensional NIG driven and Brownian motion driven cointegrated state space process.

		NIG			Brownian motion		
	True	Mean	Bias	Std. dev.	Mean	Bias	Std. dev.
ϑ_1	-1	-0.9857	-0.0143	0.0515	-0.9895	-0.0105	0.0425
ϑ_2	-2	-2.0025	0.0025	0.0573	-1.9934	-0.0066	0.0459
ϑ_3	1	0.9919	0.0081	0.0749	0.9898	0.0102	0.0570
ϑ_4	-2	-1.9758	-0.0242	0.1126	-1.9701	-0.0299	0.0872
ϑ_5	-3	-2.9774	-0.0226	0.0497	-2.9898	-0.0102	0.0324
ϑ_6	1	1.0129	-0.0129	0.1071	1.0155	-0.0155	0.0789
ϑ_7	2	2.0005	-0.0005	0.0690	2.0068	-0.0068	0.0441
ϑ_8	1	1.0078	-0.0078	0.0684	1.0096	-0.0096	0.0482
ϑ_9	1	0.9872	0.0128	0.0761	0.9777	0.0223	0.0599
ϑ_{10}	0.4751	0.4715	0.0036	0.0678	0.5200	-0.0449	0.0518
ϑ_{11}	-0.1622	-0.1572	-0.0050	0.0381	-0.1283	-0.0339	0.0266
ϑ_{12}	0.3708	0.3698	0.0010	0.0314	0.3195	0.0513	0.0213
ϑ_{13}	3	2.9999	0.0001	0.0075	2.9981	0.0019	0.0068

In Table 1 the sample mean, bias and sample standard deviation of the 350 replications of the estimated parameters are summarized. From this we see that in both the NIG driven as well the Brownian motion driven model the bias and the sample standard deviation are quite low which reflect the consistency of our estimator. Moreover, for the Brownian motion driven model the sample standard deviation is for all parameters lower than for the NIG driven model which is not surprising since the Kalman filter as well as the quasi-maximum likelihood function are motivated from the Gaussian case. In contrast, the bias in the NIG driven model is often lower than in the Gaussian model. It attracts attention that in both models the cointegration parameter ϑ_{13} has the lowest bias and sample standard deviation of all estimated parameters. This is in accordance with the fact that the consistency rate for the long-run parameters is faster than that for the short-run parameters.

Next, we investigate what happens if we use as underlying parameter space in the QML method a space which does not contain the true model. In the first parameter space Θ_I, we set $B_{2,\vartheta} = 0_{3\times 2}$ and all other matrices as above. Hence, Y_ϑ for $\vartheta \in \Theta_I$ is integrated but not cointegrated. In the second parameter space Θ_W, we set $C_{1,\vartheta} = (0,1)^T$ and all other matrices as above such that Y_ϑ for $\vartheta \in \Theta_W$ is cointegrated but the cointegration space does not model the true cointegration space. Finally, in the last parameter space Θ_S, we set $C_{1,\vartheta} = (0,0)^T$ and all other matrices as above such that Y_ϑ for $\vartheta \in \Theta_S$ is stationary and coincides with $Y_{\vartheta,st}$. The sample mean, sample standard deviation, minimal value and maximal value of the minimum of the quasi-maximum likelihood function for 100 replications of the Brownian motion driven model in the four different spaces is presented in Table 2. Of course in the space Θ, containing

Table 2

Sample mean, sample standard deviation, minimum and maximum of 100 replications of the minimum of the quasi-maximum likelihood function of the Brownian motion driven model in four different parameter spaces.

	Θ true pro.	Θ_I int. pro.	Θ_W wrong coint. pro.	Θ_S stat. pro.
Mean	5.2303	5.2851	16.2713	23.8473
St. dev.	0.0449	0.0956	11.3465	16.0159
Min	5.1226	5.1367	6.0526	9.4492
Max	5.3356	5.7509	79.3741	88.2747

the true model, the sample mean of the minimum of the likelihood function is lowest. However, the sample mean for the space Θ_I is not to far away because there at least the long-run parameter is estimated more or less appropriate such that due to Proposition 2.5 and Lemma 3.4 we get $\inf_{\vartheta\in\Theta_I} \widehat{\mathcal{L}}_n^{(h)}(\vartheta) = \inf_{\vartheta\in\Theta_I} \mathcal{L}_{n,2}^{(h)}(\vartheta_2) + o_p(1) \xrightarrow{p} \inf_{\vartheta\in\Theta_I} \boldsymbol{\mathcal{L}}_2^{(h)}(\vartheta_2)$. However, the standard deviation is much lower in Θ than in Θ_I. In contrast to the spaces Θ_W and Θ_S where the likelihood function seems to diverge. This is in accordance to the results of this paper because due to Proposition 2.5, Lemma 3.4 and (3.8) we have $\inf\limits_{\vartheta\in\Theta_W} \widehat{\mathcal{L}}_n^{(h)}(\vartheta) \xrightarrow{p} \infty$ and $\inf\limits_{\vartheta\in\Theta_S} \widehat{\mathcal{L}}_n^{(h)}(\vartheta) \xrightarrow{p} \infty$.

5.2. Three-dimensional state space model

In this example, the canonical parametrization of the cointegrated state space model has the form

$$A_{2,\vartheta} = \begin{pmatrix} \vartheta_1 & \vartheta_2 & 0 & \vartheta_3 \\ 0 & 0 & 1 & 0 \\ \vartheta_4 & \vartheta_5 & \vartheta_6 & \vartheta_7 \\ \vartheta_8 & \vartheta_9 & \vartheta_{10} & \vartheta_{11} \end{pmatrix},$$

$$B_{2,\vartheta} = \begin{pmatrix} \vartheta_1 & \vartheta_2 & \vartheta_3 \\ \vartheta_{12} & \vartheta_{13} & \vartheta_{14} \\ \vartheta_4 + \vartheta_6\vartheta_{12} & \vartheta_5 + \vartheta_6\vartheta_{13} & \vartheta_7 + \vartheta_6\vartheta_{14} \\ \vartheta_8 + \vartheta_{10}\vartheta_{12} & \vartheta_9 + \vartheta_{10}\vartheta_{13} & \vartheta_{11} + \vartheta_{10}\vartheta_{14} \end{pmatrix},$$

$$C_{2,\vartheta} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

$$B_{1,\vartheta} = \begin{pmatrix} \vartheta_{15} & \vartheta_{16} & \vartheta_{17} \\ \vartheta_{18} & \vartheta_{19} & \vartheta_{20} \end{pmatrix},$$

$$\mathrm{vech}(\Sigma^L_\vartheta) = (\vartheta_{21}, \vartheta_{22}, \vartheta_{23}, \vartheta_{24}, \vartheta_{25}, \vartheta_{26}),$$

$$C_{1,\vartheta} = \begin{pmatrix} \frac{\vartheta_{27}^2+\vartheta_{28}^2-1}{\vartheta_{27}^2+\vartheta_{28}^2+1} & 0 \\ \frac{2\cdot\vartheta_{27}}{\vartheta_{27}^2+\vartheta_{28}^2+1} & \frac{\vartheta_{28}}{\sqrt{\vartheta_{27}^2+\vartheta_{28}^2}} \\ \frac{2\cdot\vartheta_{28}}{\vartheta_{27}^2+\vartheta_{28}^2+1} & -\frac{\vartheta_{27}}{\sqrt{\vartheta_{27}^2+\vartheta_{28}^2}} \end{pmatrix}.$$

The state space model has two common stochastic trends and the cointegration space is a one-dimensional subspace of $\mathbb{R}^3$. In total we have 28 parameters. In Table 3 the sample mean, bias and sample standard deviation of the estimated parameters of 350 replications are summarized for both the NIG driven as well the Brownian motion driven model. In order to obtain the covariance matrix of the NIG Lévy process, we have had to set the parameters of the NIG Lévy process to

$$\delta = 1, \quad \alpha = 3, \quad \beta = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad \Delta = \begin{pmatrix} 1.25 & -0.5 & \frac{1}{6}\sqrt{3} \\ -0.5 & 1 & -\frac{1}{3}\sqrt{3} \\ \frac{1}{6}\sqrt{3} & -\frac{1}{3}\sqrt{3} & \frac{4}{3} \end{pmatrix} \quad \text{and}$$

$$\mu = -\frac{1}{2\sqrt{31}} \begin{pmatrix} 3 \\ 2 \end{pmatrix}.$$

The results are very similar to the two-dimensional example. In most cases the sample standard deviation in the Brownian motion driven model is lower than in the NIG driven model. Moreover, the bias and the standard deviation of the long-run parameters $(\vartheta_{27}, \vartheta_{28})$ are lower than the values of the other parameters.

6. Conclusion

The main contribution of the present paper is the development of a QML estimation procedure for the parameters of cointegrated solutions of continuous-time linear state space models sampled equidistantly allowing flexible margins. We showed that the QML estimator for the long-run parameter is super-consistent and that of the short-run parameter is consistent. Moreover, the QML estimator for the long-run parameter converges with a n-rate to a mixed normal distribution, whereas the short-run parameter converges with a $\sqrt{n}$-rate to a normal distribution. In the simulation study, we saw that the estimator works quite well in practice.

Table 3
Sample mean, bias and sample standard deviation of 350 replications of the QML estimator for a three-dimensional NIG driven and Brownian motion driven cointegrated state space process.

		NIG			Brownian motion		
	True	Mean	Bias	Std. dev.	Mean	Bias	Std. dev.
ϑ_1	-2	-1.9910	-0.0090	0.0583	-1.9958	-0.0042	0.0475
ϑ_2	-3	-3.0042	0.0042	0.0407	-3.0005	0.0005	0.0339
ϑ_3	-3	-3.0194	0.0194	0.0456	-3.0309	0.0309	0.0401
ϑ_4	1	0.9887	0.0113	0.0440	0.9987	0.0013	0.0381
ϑ_5	1	0.9977	0.0023	0.0351	0.9895	0.0105	0.0316
ϑ_6	-1	-0.9861	-0.0139	0.0544	-0.9763	-0.0237	0.0431
ϑ_7	2	2.0122	-0.0122	0.0396	2.0113	-0.0113	0.0342
ϑ_8	-1	-1.0039	0.0039	0.0442	-1.0075	0.0075	0.0399
ϑ_9	-3	-2.9937	-0.0063	0.0342	-2.9896	-0.0104	0.0348
ϑ_{10}	-3	-2.9904	-0.0096	0.0490	-2.9892	-0.0108	0.0444
ϑ_{11}	-1	-1.0055	0.0055	0.0449	-1.0097	0.0097	0.0461
ϑ_{12}	-1	-1.0023	0.0023	0.0386	-1.0242	0.0242	0.0367
ϑ_{13}	2	1.9984	0.0016	0.0363	2.0077	-0.0077	0.0295
ϑ_{14}	1	1.0034	-0.0034	0.0353	0.9740	0.0260	0.0353
ϑ_{15}	1	0.9984	0.0016	0.0351	1.0175	-0.0175	0.0284
ϑ_{16}	0	-0.0345	0.0345	0.0644	-0.0361	0.0361	0.0513
ϑ_{17}	1	0.9840	0.0160	0.0521	0.9623	0.0377	0.0417
ϑ_{18}	1	1.0010	-0.0010	0.0314	0.9877	0.0123	0.0303
ϑ_{19}	-2	-1.9841	-0.0159	0.0388	-1.9868	-0.0132	0.0306
ϑ_{20}	0	0.0111	-0.0111	0.0347	-0.0090	0.0090	0.0362
ϑ_{21}	0.5310	0.5279	0.0031	0.0605	0.5849	-0.0539	0.0478
ϑ_{22}	-0.1934	-0.1870	-0.0064	0.0385	-0.2037	0.0103	0.0328
ϑ_{23}	0.1678	0.1678	0.0000	0.0467	0.1513	0.0165	0.0396
ϑ_{24}	0.3784	0.3816	-0.0032	0.0293	0.4209	-0.0425	0.0259
ϑ_{25}	-0.2227	-0.2127	0.0100	0.0334	-0.2209	-0.0018	0.0300
ϑ_{26}	0.5632	0.5585	0.0047	0.0476	0.4814	0.0818	0.0356
ϑ_{27}	1	1.0002	0.0002	0.0030	0.9995	0.0005	0.0033
ϑ_{28}	2	2.0000	0.0000	0.0079	2.0004	-0.0004	0.0091

In this paper, we lay the mathematical basis for QML estimation for cointegrated solutions of linear state-space models. In a separate paper Fasen-Hartmann and Scholz [19] we present an algorithm to construct canonical forms for the state space model satisfying the assumptions of this paper, which is necessary to apply the method to data. We decided to split the paper because the introduction into a canonical form is quite lengthy and would blow up the present paper. Moreover, a drawback of our estimation procedure is that we assume that the cointegration rank is known in advance which is not the case in reality. First, we have to estimate and test the cointegration rank. For this it is possible to incorporate some well-known results for estimating and testing the cointegration rank of cointegrated VARMA processes as, e.g., presented in Bauer and Wagner [5], Lütkepohl and Claessen [36], Saikkonen [45], Yap and Reinsel [59]. This will also be considered in Fasen-Hartmann and Scholz [19]. Some parts of Fasen-Hartmann and Scholz [19] can already be found in Scholz [52].

Appendix A: Auxiliary results

A.1. Asymptotic results

For the derivation of the asymptotic behavior of our estimators we require the asymptotic behavior of the standardized score vector and the standardized Hessian matrix. To obtain these asymptotic results we use the next proposition.

Proposition A.1. *Let Assumption A hold. Furthermore, let $(L_k^{(h)})_{k\in\mathbb{N}_0} := (L(kh))_{k\in\mathbb{N}_0}$ be the Lévy process sampled at distance h and $\Delta L_k^{(h)} = L(kh) - L((k-1)h)$. Define for $n \in \mathbb{N}, k \in \mathbb{N}_0$,*

$$\xi_k^{(h)} := \begin{pmatrix} \Delta Y_k^{(h)} \\ Y_{st,k}^{(h)} \\ \Delta L_k^{(h)} \end{pmatrix} \quad \text{and} \quad S_n^{(h)} := \sum_{k=1}^n \xi_k^{(h)}.$$

Let $\bar{\mathsf{l}}(z,\vartheta) := \sum_{i=0}^\infty \bar{\mathsf{l}}_i(\vartheta) z^i$ and $\underline{\mathsf{l}}(z,\vartheta) := \sum_{i=0}^\infty \underline{\mathsf{l}}_i(\vartheta) z^i$, $\vartheta \in \Theta$, $z \in \mathbb{C}$, where $(\bar{\mathsf{l}}_i(\vartheta))_{i\in\mathbb{N}_0}$ is a deterministic uniformly exponentially bounded continuous matrix sequence in $\mathbb{R}^{\overline{d}\times(2d+m)}$ and similarly $(\underline{\mathsf{l}}_i(\vartheta))_{i\in\mathbb{N}_0}$ is a nonstochastic uniformly exponentially bounded continuous matrix sequence in $\mathbb{R}^{\underline{d}\times(2d+m)}$. Moreover, $\overline{\Pi}(\vartheta) \in \mathbb{R}^{\overline{d}\times(2d+m)}$, $\underline{\Pi}(\vartheta) \in \mathbb{R}^{\underline{d}\times(2d+m)}$ are continuous matrix functions as well. We write $\underline{\mathsf{l}}(\mathsf{B},\vartheta)\xi^{(h)} = (\underline{\mathsf{l}}(\mathsf{B},\vartheta)\xi_k^{(h)})_{k\in\mathbb{N}_0}$ with $\underline{\mathsf{l}}(\mathsf{B},\vartheta)\xi_k^{(h)} = \sum_{i=0}^\infty \bar{\mathsf{l}}_i(\vartheta)\xi_{k-i}^{(h)}$ and similarly $\bar{\mathsf{l}}(\mathsf{B},\vartheta)\xi^{(h)}$. Let $(W(r))_{0\le r\le 1} = ((W_1(r)^\mathsf{T}, W_2(r)^\mathsf{T}, W_3(r)^\mathsf{T})^\mathsf{T})_{0\le r\le 1}$ be the Brownian motion as defined on p. 5166. Then, the following statements hold for $j \in \mathbb{N}_0$.

(a) $$\sup_{\vartheta\in\Theta}\left\| \frac{1}{n}\sum_{k=1}^n [\bar{\mathsf{l}}(\mathsf{B},\vartheta)\xi_k^{(h)}][\underline{\mathsf{l}}(\mathsf{B},\vartheta)\xi_{k+j}^{(h)}]^\mathsf{T} - \mathbb{E}\left[[\bar{\mathsf{l}}(\mathsf{B},\vartheta)\xi_1^{(h)}][\underline{\mathsf{l}}(\mathsf{B},\vartheta)\xi_{1+j}^{(h)}]^\mathsf{T}\right]\right\| \xrightarrow{p} 0.$$

(b) $$n^{-2}\sum_{k=1}^n \overline{\Pi}(\vartheta) S_{k-1}^{(h)}[S_{k-1}^{(h)}]^\mathsf{T}\underline{\Pi}(\vartheta)^\mathsf{T} \xrightarrow{w} \overline{\Pi}(\vartheta)\int_0^1 W(r)W(r)^\mathsf{T}\,\mathrm{d}r\,\underline{\Pi}(\vartheta)^\mathsf{T}.$$

(c) $$n^{-1}\sum_{k=1}^n \overline{\Pi}(\vartheta) S_{k-1}^{(h)}[\underline{\mathsf{l}}(\mathsf{B},\vartheta)\xi_k^{(h)}]^\mathsf{T} \xrightarrow{w} \overline{\Pi}(\vartheta)\int_0^1 W(r)\mathrm{d}W(r)^\mathsf{T}\underline{\mathsf{l}}(1,\vartheta)^\mathsf{T} + \sum_{j=1}^\infty \Gamma_{\overline{\Pi}(\vartheta)\xi^{(h)},\underline{\mathsf{l}}(\mathsf{B},\vartheta)\xi^{(h)}}(j).$$

The stated weak convergence results also hold jointly.

Before we state the proof of Proposition A.1 we need some auxiliary results.

Lemma A.2. *Let ψ be defined as in (3.2). Then, the following statements hold for $l \in \mathbb{N}_0$.*

(a) $\mathbb{E}(\xi_k^{(h)}) = 0_{2d+m}$ *and* $\mathbb{E}\|\xi_k^{(h)}\|^4 < \infty$.

(b) $\frac{1}{n}\sum_{k=1}^n \xi_k^{(h)} \xrightarrow{p} 0_{2d+m}$, $\frac{1}{n}\sum_{k=1}^n \xi_k^{(h)}[\xi_{k+l}^{(h)}]^\mathsf{T} \xrightarrow{p} \mathbb{E}(\xi_1^{(h)}[\xi_{1+l}^{(h)}]^\mathsf{T}) =: \Gamma_{\xi^{(h)}}(l)$.

(c) $\sum_{l=0}^{\infty} \mathbb{E}\|\xi_1^{(h)}[\xi_{1+l}^{(h)}]^\mathsf{T}\| < \infty$.

(d) $\left(\frac{1}{\sqrt{n}} S_{\lfloor nr \rfloor}^{(h)}\right)_{0\leq r\leq 1} \xrightarrow{w} (\psi(1)W^*(r))_{0\leq r\leq 1}$ *where* $(W^*(r))_{0\leq r\leq 1}$ *is a* $(m+(N-c))$*-dimensional Brownian motion with covariance matrix*

$$\Sigma_{W^*} = \int_0^h \begin{pmatrix} \Sigma_L & \Sigma_L B_2^T \mathrm{e}^{A_2^T u} \\ \mathrm{e}^{A_2 u} B_2 \Sigma_L & \mathrm{e}^{A_2 u} B_2 \Sigma_L B_2^T \mathrm{e}^{A_2^T u} \end{pmatrix} du$$

and ψ *is defined as in* (3.2).

(e) $\frac{1}{n}\sum_{k=2}^{n} S_{k-1}^{(h)}[\xi_k^{(h)}]^\mathsf{T} \xrightarrow{w} \psi(1)\int_0^1 W^*(r)\mathrm{d}W^*(r)^\mathsf{T}\psi(1)^\mathsf{T} + \sum_{l=1}^{\infty}\Gamma_{\xi^{(h)}}(l)$.

Proof. We shortly sketch the proof. The sequence $(\xi_k^{(h)})_{k\in\mathbb{N}}$ has the MA representation

$$\begin{pmatrix} \Delta Y_k^{(h)} \\ Y_{st,k}^{(h)} \\ \Delta L_k^{(h)} \end{pmatrix} = \xi_k^{(h)} = \sum_{j=-\infty}^{k} \psi_{k-j}\eta_j^{(h)} \tag{A.1}$$

with the iid sequence

$$\eta_k^{(h)} := \left(\Delta L_k^{(h)\mathsf{T}},\ R_k^{(h)\mathsf{T}} \right)^\mathsf{T} \quad \text{and} \quad R_k^{(h)} := \int_{(k-1)h}^{kh} \mathrm{e}^{A_2(kh-u)} B_2 \,\mathrm{d}L_u.$$

Hence, $(\xi_k^{(h)})_{k\in\mathbb{N}}$ is stationary and ergodic as a measurable map of a stationary ergodic process (see Krengel [34, Theorem 4.3 in Chapter 1]).

(a) is due to Assumption A.

(b) is a direct consequence of Birkhoff's ergodic theorem.

(c) follows from $\mathbb{E}\|\eta_k^{(h)}\|^4 < \infty$, $\|\psi_j\| \leq \mathfrak{C}\rho^j$ for some $\mathfrak{C} > 0$, $0 < \rho < 1$ and the MA-representation (A.1).

(d, e) are conclusions of Johansen [31, Theorem B.13] and the MA-representation (A.1). □

Proof of Proposition A.1. (a) The proof follows directly by Theorem 4.1 of Saikkonen [46] and the comment of Saikkonen [46, p.163, line 4] if we can show that Assumption 4.1 and 4.2 of that paper are satisfied. Since we have uniformly exponentially bounded families of matrix sequences, Saikkonen [46, Assumption 4.1] is obviously satisfied. Saikkonen [46, Assumption 4.2] is satisfied due to Lemma A.2.

Note that we have two different coefficient matrices, whereas the results in Saikkonen [46] are proved for the same coefficient matrix. However, Saikkonen [46, Theorem 4.1] also holds if the coefficient matrices are different as long as each sequence of matrix coefficients satisfies the necessary conditions as mentioned in the paper of Saikkonen [46, p. 163].

(b, c) Due to Lemma A.2, Saikkonen [46, Assumption 4.3] is satisfied as well. Hence, we can conclude the weak convergence result from Saikkonen [46, Theorem 4.2(iii)] and [46, Theorem 4.2(iv)], respectively. □

A.2. Lipschitz continuity results

A kind of local Lipschitz continuity in ϑ^0 for the processes in Proposition A.1 is presented next. The local Lipschitz continuity in ϑ^0 implies, in particular, local stochastic equicontinuity in ϑ^0. However, this kind of local Lipschitz continuity in ϑ^0 is stronger than local stochastic equicontinuity in ϑ^0 so that we are not able to apply the stochastic equicontinuity results of Saikkonen [46, 47] directly. The stochastic equicontinuity of the process in Proposition A.3(a) and (c) can be deduced with some effort from Saikkonen [46, 47] but the process in Proposition A.3(b) is not covered in these papers.

Proposition A.3. *Let the assumption and notation of Proposition A.1 hold. Assume further that $\overline{\Pi}(\vartheta)$, $\underline{\Pi}(\vartheta)$ are Lipschitz-continuous and the sequence of matrix functions $(\nabla_\vartheta(\bar{\mathsf{l}}_i(\vartheta)))_{i\in\mathbb{N}_0}$ and $(\nabla_\vartheta(\underline{\mathsf{l}}_i(\vartheta)))_{i\in\mathbb{N}_0}$ are uniformly exponentially bounded.*

(a) Define $\mathcal{X}_n(\vartheta) = \sum_{k=1}^n \overline{\Pi}(\vartheta) S_{k-1}^{(h)} [S_{k-1}^{(h)}]^\mathsf{T} \underline{\Pi}(\vartheta)^\mathsf{T}$. Then,

$$\|\mathcal{X}_n(\vartheta) - \mathcal{X}_n(\vartheta^0)\| \leq \mathfrak{C}\|\vartheta - \vartheta^0\| \left\| \sum_{k=1}^n S_{k-1}^{(h)} [S_{k-1}^{(h)}]^\mathsf{T} \right\|. \tag{A.2}$$

If additionally $\underline{\Pi}(\vartheta^0) = 0_{\underline{d}\times(2d+m)}$ and $\overline{\Pi}(\vartheta^0) = 0_{\overline{d}\times(2d+m)}$ then

$$\|\mathcal{X}_n(\vartheta) - \mathcal{X}_n(\vartheta^0)\| \leq \mathfrak{C}\|\vartheta - \vartheta^0\|^2 \left\| \sum_{k=1}^n S_{k-1}^{(h)} [S_{k-1}^{(h)}]^\mathsf{T} \right\|. \tag{A.3}$$

(b) Define $\mathcal{X}_n(\vartheta) = \sum_{k=1}^n \overline{\Pi}(\vartheta) S_{k-1}^{(h)} [\underline{\mathsf{l}}(\mathsf{B},\vartheta)\xi_k^{(h)}]^\mathsf{T}$. Then,

$$\|\mathcal{X}_n(\vartheta) - \mathcal{X}_n(\vartheta^0)\| \leq \mathfrak{C} n \|\vartheta - \vartheta^0\| V_n \tag{A.4}$$

where

$$\begin{aligned} V_n &= \left\| \frac{1}{n}\sum_{k=1}^n S_{k-1}^{(h)} [\xi_k^{(h)}]^T \right\| + \|S_n^{(h)}\| [\mathsf{k}_\rho(\mathsf{B})\|\xi_n^{(h)}\|]] \\ &\quad + \frac{1}{n}\sum_{k=1}^n \|\Delta S_k^{(h)}\| [\mathsf{k}_\rho(\mathsf{B})\|\xi_k^{(h)}\| + \left\| \frac{1}{n}\sum_{k=1}^n S_{k-1}^{(h)} \left[\underline{\mathsf{l}}(\mathsf{B},\vartheta^0)\xi_k^{(h)}\right]^\mathsf{T} \right\| \\ &= \mathcal{O}_p(1), \end{aligned} \tag{A.5}$$

$\mathsf{k}_\rho(z) = \mathfrak{c}\sum_{j=0}^\infty \rho^j z^j$ *for some* $0 < \rho < 1$, $\mathfrak{c} > 0$, *and* $\mathsf{k}_\rho(\mathsf{B})\|\xi_k^{(h)}\| := \mathfrak{c}\sum_{j=0}^\infty \rho^j \|\xi_{k-j}^{(h)}\|$.

(c) *Define* $\mathcal{X}_n(\vartheta) = \sum_{k=1}^n [\bar{\mathsf{l}}(\mathsf{B},\vartheta)\xi_k^{(h)}][\underline{\mathsf{l}}(\mathsf{B},\vartheta)\xi_k^{(h)}]^\mathsf{T}$. *Then, there exists a random variable* $Q_n = \mathcal{O}_p(1)$ *so that*

$$\|\mathcal{X}_n(\vartheta) - \mathcal{X}_n(\vartheta^0)\| \leq \mathfrak{C} n\|\vartheta - \vartheta^0\| Q_n. \tag{A.6}$$

In particular, $V_n + Q_n = \mathcal{O}_p(1)$.

Proof.

(a) We have the upper bound

$$\begin{aligned}
&\|\mathcal{X}_n(\vartheta) - \mathcal{X}_n(\vartheta^0)\| \\
&\quad\leq \left\|\sum_{k=1}^n [\overline{\Pi}(\vartheta) - \overline{\Pi}(\vartheta^0)] S_{k-1}^{(h)}[S_{k-1}^{(h)}]^\mathsf{T}\underline{\Pi}(\vartheta)^\mathsf{T}\right\| \\
&\qquad + \left\|\sum_{k=1}^n \overline{\Pi}(\vartheta^0) S_{k-1}^{(h)}[S_{k-1}^{(h)}]^\mathsf{T}[\underline{\Pi}(\vartheta) - \underline{\Pi}(\vartheta^0)]^\mathsf{T}\right\| \\
&\quad\leq \left(\|\overline{\Pi}(\vartheta) - \overline{\Pi}(\vartheta^0)\|\|\underline{\Pi}(\vartheta)\| + \|\underline{\Pi}(\vartheta) - \underline{\Pi}(\vartheta^0)\|\|\overline{\Pi}(\vartheta^0)\|\right)\left\|\sum_{k=1}^n S_{k-1}^{(h)}[S_{k-1}^{(h)}]^\mathsf{T}\right\|.
\end{aligned}$$

Since $\overline{\Pi}(\vartheta)$ and $\underline{\Pi}(\vartheta)$ are Lipschitz continuous,

$$\max(\|\overline{\Pi}(\vartheta) - \overline{\Pi}(\vartheta^0)\|, \|\underline{\Pi}(\vartheta) - \underline{\Pi}(\vartheta^0)\|) \leq \mathfrak{C}\|\vartheta - \vartheta^0\|$$

and $\sup_{\vartheta\in\Theta}\max(\|\overline{\Pi}(\vartheta)\|, \|\underline{\Pi}(\vartheta)\|) \leq \mathfrak{C}$. Thus, (A.2) holds. Moreover, (A.3) is valid because for $\underline{\Pi}(\vartheta^0) = 0_{\underline{d},2d+m}$, $\overline{\Pi}(\vartheta^0) = 0_{\overline{d},2d+m}$ the upper bound

$$\begin{aligned}
\|\mathcal{X}_n(\vartheta) - \mathcal{X}_n(\vartheta^0)\| &= \|\mathcal{X}_n(\vartheta)\| \\
&\leq \|\underline{\Pi}(\vartheta)\|\|\overline{\Pi}(\vartheta)\|\left\|\sum_{k=1}^n S_{k-1}^{(h)}[S_{k-1}^{(h)}]^\mathsf{T}\right\| \\
&\leq \|\underline{\Pi}(\vartheta) - \underline{\Pi}(\vartheta^0)\|\|\overline{\Pi}(\vartheta) - \overline{\Pi}(\vartheta^0)\|\left\|\sum_{k=1}^n S_{k-1}^{(h)}[S_{k-1}^{(h)}]^\mathsf{T}\right\| \\
&\leq \mathfrak{C}\|\vartheta - \vartheta^0\|^2\left\|\sum_{k=1}^n S_{k-1}^{(h)}[S_{k-1}^{(h)}]^\mathsf{T}\right\|
\end{aligned}$$

is valid.

(b) Using a Taylor expansion leads to

$$\mathrm{vec}(\underline{\mathsf{l}}(z,\vartheta)) - \mathrm{vec}(\underline{\mathsf{l}}(z,\vartheta^0)) = \sum_{j=0}^\infty [\mathrm{vec}(\underline{\mathsf{l}}_j(\vartheta)) - \mathrm{vec}(\underline{\mathsf{l}}_j(\vartheta^0))]z^j$$

$$= \sum_{j=0}^{\infty} \nabla_{\vartheta} \mathrm{vec}(\underline{\mathsf{l}}_j^*(\underline{\vartheta(j)}))(\vartheta - \vartheta^0) z^j,$$

where $\mathrm{vec}(\underline{\mathsf{l}}_j^*(\underline{\vartheta(j)}))$ denotes the matrix whose i^{th} row is equal to the i^{th} row of $\mathrm{vec}(\underline{\mathsf{l}}_j(\underline{\vartheta^i(j)}))$ with $\underline{\vartheta^i(j)} \in \Theta$ such that $\|\underline{\vartheta^i(j)} - \vartheta^0\| \leq \|\vartheta - \vartheta^0\|$ for $i = 1, \ldots, \underline{d}(2d+m)$. Due to assumption, $\|\nabla_{\vartheta} \mathrm{vec}(\underline{\mathsf{l}}_j^*(\underline{\vartheta(j)}))\| \leq \mathfrak{C}\rho^j$ for $j \in \mathbb{N}_0$ and some $0 < \rho < 1$ so that

$$\|\underline{\mathsf{l}}(z,\vartheta) - \underline{\mathsf{l}}(z,\vartheta^0)\| \leq \mathsf{k}_{\rho}(|z|)\|\vartheta - \vartheta^0\|. \tag{A.7}$$

Hence,

$$\| \left(\underline{\mathsf{l}}(\mathsf{B},\vartheta) - \underline{\mathsf{l}}(\mathsf{B},\vartheta^0)\right) \xi_k^{(h)} \| \leq \|\vartheta - \vartheta^0\| \left[\mathsf{k}_{\rho}(\mathsf{B})\|\xi_k^{(h)}\|\right]. \tag{A.8}$$

Define $\underline{\mathsf{l}}^{\nabla}(z,\vartheta,\vartheta^0) := \underline{\mathsf{l}}(z,\vartheta) - \underline{\mathsf{l}}(z,\vartheta^0) =: \sum_{j=0}^{\infty} \underline{\mathsf{l}}_j^{\nabla}(\vartheta,\vartheta^0) z^j$. Then, we apply a Beveridge-Nelson decomposition (see Saikkonen [46, (9)]) to get

$$\underline{\mathsf{l}}^{\nabla}(\mathsf{B},\vartheta,\vartheta^0)\xi_k^{(h)} = \underline{\mathsf{l}}^{\nabla}(1,\vartheta,\vartheta^0)\xi_k^{(h)} + \eta_k(\vartheta,\vartheta^0) - \eta_{k-1}(\vartheta,\vartheta^0)$$

with $\eta_k(\vartheta,\vartheta^0) := -\sum_{j=0}^{\infty}\sum_{i=j+1}^{\infty} \underline{\mathsf{l}}_i^{\nabla}(\vartheta,\vartheta^0)\xi_{k-j}^{(h)}$. Thus,

$$\begin{aligned}
&\frac{1}{n}\sum_{k=1}^{n} \overline{\Pi}(\vartheta) S_{k-1}^{(h)} \left[\left(\underline{\mathsf{l}}(\mathsf{B},\vartheta) - \underline{\mathsf{l}}(\mathsf{B},\vartheta^0)\right)\xi_k^{(h)}\right]^{\mathsf{T}} \\
&\quad = \overline{\Pi}(\vartheta)\frac{1}{n}\sum_{k=1}^{n} S_{k-1}^{(h)}[\xi_k^{(h)}]^{\mathsf{T}}\underline{\mathsf{l}}^{\nabla}(1,\vartheta,\vartheta^0)^{\mathsf{T}} + \overline{\Pi}(\vartheta)\frac{1}{n}\sum_{k=1}^{n} S_{k-1}^{(h)}[\eta_k(\vartheta,\vartheta^0)]^{\mathsf{T}} \\
&\quad\quad -\overline{\Pi}(\vartheta)\frac{1}{n}\sum_{k=1}^{n} S_{k-1}^{(h)}[\eta_{k-1}(\vartheta,\vartheta^0)]^{\mathsf{T}} \\
&\quad = \overline{\Pi}(\vartheta)\frac{1}{n}\sum_{k=1}^{n} S_{k-1}^{(h)}[\xi_k^{(h)}]^{\mathsf{T}}\underline{\mathsf{l}}^{\nabla}(1,\vartheta,\vartheta^0)^{\mathsf{T}} + \overline{\Pi}(\vartheta) S_n^{(h)}[\eta_n(\vartheta,\vartheta^0)]^{T} \\
&\quad\quad -\overline{\Pi}(\vartheta)\frac{1}{n}\sum_{k=1}^{n} \Delta S_k^{(h)}[\eta_k(\vartheta,\vartheta^0)]^{\mathsf{T}}.
\end{aligned}$$

Due to (A.7), $\|\underline{\mathsf{l}}^{\nabla}(1,\vartheta,\vartheta^0)\| \leq \mathfrak{C}\|\vartheta - \vartheta^0\|$ and $\|\underline{\mathsf{l}}_j^{\nabla}(\vartheta,\vartheta^0)\| \leq \mathfrak{C}\|\vartheta - \vartheta^0\|\rho^j$ so that $\|\eta_k(\vartheta,\vartheta^0)\| \leq \|\vartheta - \vartheta^0\| \left[\mathsf{k}_{\rho}(\mathsf{B})\|\xi_k^{(h)}\|\right]$ as well. Finally, we receive

$$\begin{aligned}
&\left\|\overline{\Pi}(\vartheta)\frac{1}{n}\sum_{k=1}^{n} S_{k-1}^{(h)} \left[\left(\underline{\mathsf{l}}(\mathsf{B},\vartheta) - \underline{\mathsf{l}}(\mathsf{B},\vartheta^0)\right)\xi_k^{(h)}\right]^{\mathsf{T}}\right\| \\
&\quad \leq \mathfrak{C}\|\vartheta - \vartheta^0\| \left[\left\|\frac{1}{n}\sum_{k=1}^{n} S_{k-1}^{(h)}[\xi_k^{(h)}]^{\mathsf{T}}\right\|\right. \\
&\quad\quad \left. + \|S_n^{(h)}\| \left[\mathsf{k}_{\rho}(\mathsf{B})\|\xi_n^{(h)}\|\right] + \frac{1}{n}\sum_{k=1}^{n} \|\Delta S_k^{(h)}\| \left[\mathsf{k}_{\rho}(\mathsf{B})\|\xi_k^{(h)}\|\right]\right]
\end{aligned} \tag{A.9}$$

and

$$\left\|[\overline{\Pi}(\vartheta)-\overline{\Pi}(\vartheta^0)]\frac{1}{n}\sum_{k=1}^{n}S_{k-1}^{(h)}\left[\underline{\mathsf{l}}(\mathsf{B},\vartheta^0)\xi_k^{(h)}\right]^{\mathsf{T}}\right\|$$

$$\leq \mathfrak{C}\|\vartheta-\vartheta^0\|\left\|\frac{1}{n}\sum_{k=1}^{n}S_{k-1}^{(h)}\left[\underline{\mathsf{l}}(\mathsf{B},\vartheta^0)\xi_k^{(h)}\right]^{\mathsf{T}}\right\|. \tag{A.10}$$

Then, (A.9) and (A.10) result in (A.4).

It remains to prove (A.5). The first term $\frac{1}{n}\sum_{k=1}^{n}S_{k-1}^{(h)}[\xi_k^{(h)}]^T$ in the definition of V_n is $\mathcal{O}_p(1)$ by Lemma A.2. Moreover, $\frac{1}{n}S_n^{(h)}=\frac{1}{n}\sum_{k=1}^{n}\xi_k^{(h)}=\mathcal{O}_p(1)$ by Birkhoff's Ergodic Theorem; similarly the third term $\frac{1}{n}\sum_{k=1}^{n}\|\Delta S_k^{(h)}\|\times\left[\mathsf{k}_\rho(\mathsf{B})\|\xi_k^{(h)}\|\right]$ is $\mathcal{O}_p(1)$ by Birkhoff's Ergodic Theorem. Finally, the last term $\frac{1}{n}\sum\limits_{k=1}^{n}S_{k-1}^{(h)}\left[\underline{\mathsf{l}}(\mathsf{B},\vartheta^0)\xi_k^{(h)}\right]^{\mathsf{T}}$ is $\mathcal{O}_p(1)$ by Proposition A.1(c).

(c) The proof is similarly to the proof of (b). □

Appendix B: Properties of the pseudo-innovations

In this section we present some probabilistic properties of the pseudo-innovations. Therefore, we start with an auxiliary lemma on the functions defining the pseudo-innovations and the prediction covariance matrix which we require for the pseudo-innovations to be partial differentiable.

Lemma B.1. *Let Assumption A hold.*

(a) The matrix functions $\Pi(\vartheta)$, $\mathsf{k}(z,\vartheta)$, $V_\vartheta^{(h)}$ and $(V_\vartheta^{(h)})^{-1}$ are Lipschitz continuous on Θ and three times partial differentiable.
(b) $\sup_{\vartheta\in\Theta}\|(V_\vartheta^{(h)})^{-1}\|\leq\mathfrak{C}$.
(c) $\inf_{\vartheta\in\Theta}\sigma_{\min}((V_\vartheta^{(h)})^{-1})>0$.

Proof. (a) is a consequence of Assumption A and Scholz [52, Lemma 5.9.3]. However, Scholz [52, Lemma 5.9.3] shows only the twice continuous differentiability but the proof of the existence of the third partial differential is analog.
(b) follows from (a) and the compactness of Θ.
(c) Due to Scholz [52, Lemma 5.9.1] the matrix $(V_\vartheta^{(h)})^{-1}$ is non-singular. Hence, we can conclude the statement from (a) and the compactness of Θ. □

Thus, the pseudo-innovations are three times differentiable and we receive an analog version of Lemma 2.3.

Lemma B.2. *Let Assumption A hold and let $u,v\in\{1,\dots,s\}$. Then, the following results are valid.*

(a) The matrix sequence $(\partial_v\mathsf{k}_j(\vartheta))_{j\in\mathbb{N}}$ in $\mathbb{R}^{d\times d}$ is uniformly exponentially bounded such that $\partial_v\varepsilon_k^{(h)}(\vartheta)=-\partial_v\Pi(\vartheta)^{\mathsf{T}}Y_{k-1}^{(h)}-\sum_{j=1}^{\infty}\partial_v\mathsf{k}_j(\vartheta)\Delta Y_{k-j}^{(h)}$.

(b) *The matrix sequence* $(\partial_{u,v}\mathsf{k}_j(\vartheta))_{j\in\mathbb{N}}$ *in* $\mathbb{R}^{d\times d}$ *is uniformly exponentially bounded such that* $\partial_{u,v}\varepsilon_k^{(h)}(\vartheta) = -\partial_{u,v}\Pi(\vartheta)^\mathsf{T} Y_{k-1}^{(h)} - \sum_{j=1}^\infty \partial_{u,v}\mathsf{k}_j(\vartheta)\Delta Y_{k-j}^{(h)}$.

Proof. Recall the representation given in Lemma 2.3 where $(\mathsf{k}_j(\vartheta))_{j\in\mathbb{N}}$ is an uniformly exponentially bounded matrix sequence. Then, the proof is analog to Schlemm and Stelzer [51, Lemma 2.11]. □

Lemma B.3. *Let Assumption A hold and* $i,j\in\{1,\ldots,s_2\}$. *Then, the following results are valid.*

(a) $(\varepsilon_k^{(h)}(\vartheta^0)^\mathsf{T}, \partial_j^{st}\varepsilon_k^{(h)}(\vartheta^0)^\mathsf{T}, \partial_{i,j}^{st}\varepsilon_k^{(h)}(\vartheta^0)^\mathsf{T})_{k\in\mathbb{N}}$ *is a stationary and ergodic sequence.*

(b) $\mathbb{E}\|\varepsilon_k^{(h)}(\vartheta^0)\|^4<\infty$, $\mathbb{E}\|\partial_j^{st}\varepsilon_k^{(h)}(\vartheta^0)\|^4<\infty$ *and* $\mathbb{E}\|\partial_{i,j}^{st}\varepsilon_k^{(h)}(\vartheta^0)\|^4<\infty$.

(c) $\mathbb{E}(\partial_i^{st}\varepsilon_k^{(h)}(\vartheta^0)\varepsilon_k^{(h)}(\vartheta^0)^\mathsf{T}) = 0_{d\times d}$ *and* $\mathbb{E}(\partial_{i,j}^{st}\varepsilon_k^{(h)}(\vartheta^0)\varepsilon_k^{(h)}(\vartheta^0)^\mathsf{T}) = 0_{d\times d}$.

(d) $\mathbb{E}(\varepsilon_k^{(h)}(\vartheta^0)\varepsilon_k^{(h)}(\vartheta^0)^\mathsf{T}) = V_{\vartheta_0}^{(h)}$.

Proof.

(a) Representation (1.2) and Lemma B.1 yield

$$\begin{aligned}
\Pi(\vartheta^0)Y_k^{(h)} &= \alpha(\vartheta_1^0,\vartheta_2^0)(C_{1,\vartheta_1^0}^\perp)^\mathsf{T} Y_k^{(h)} = \Pi(\vartheta^0)Y_{st,k}^{(h)},\\
\partial_i^{st}\Pi(\vartheta^0)Y_k^{(h)} &= \left(\partial_i^{st}\alpha(\vartheta^0)\right)(C_{1,\vartheta_1^0}^\perp)^\mathsf{T} Y_{st,k}^{(h)},\\
\partial_{i,j}^{st}\Pi(\vartheta^0)Y_k^{(h)} &= \left(\partial_{i,j}^{st}\alpha(\vartheta^0)\right)(C_{1,\vartheta_1^0}^\perp)^\mathsf{T} Y_{st,k}^{(h)},
\end{aligned}$$

and hence,

$$\begin{aligned}
\varepsilon_k^{(h)}(\vartheta^0) &= -\Pi(\vartheta^0)Y_{st,k-1}^{(h)} + \mathsf{k}(\mathsf{B},\vartheta)\Delta Y_k^{(h)},\\
\partial_i^{st}\varepsilon_k^{(h)}(\vartheta^0) &= -\partial_i^{st}\Pi(\vartheta^0)^\mathsf{T} Y_{st,k-1}^{(h)} - \sum_{l=1}^\infty \partial_i^{st}\mathsf{k}_l(\vartheta^0)\Delta Y_{k-l}^{(h)}, \qquad \text{(B.1)}\\
\partial_{i,j}^{st}\varepsilon_k^{(h)}(\vartheta^0) &= -\partial_{i,j}^{st}\Pi(\vartheta^0)^\mathsf{T} Y_{st,k-1}^{(h)} - \sum_{l=1}^\infty \partial_{i,j}^{st}\mathsf{k}_l(\vartheta^0)\Delta Y_{k-l}^{(h)}.
\end{aligned}$$

These are obviously stationary processes. Fasen-Hartmann and Scholz [18, Proposition 5.9] state already that $(\varepsilon_k^{(h)}(\vartheta^0))_{k\in\mathbb{N}}$ is ergodic with finite second moments. The same arguments lead to the ergodicity of

$$(\varepsilon_k^{(h)}(\vartheta^0)^\mathsf{T}, \partial_j^{st}\varepsilon_k^{(h)}(\vartheta^0)^\mathsf{T}, \partial_{i,j}^{st}\varepsilon_k^{(h)}(\vartheta^0)^\mathsf{T})_{k\in\mathbb{N}}.$$

(b) The finite fourth moment of $(\varepsilon_k^{(h)}(\vartheta^0))_{k\in\mathbb{N}}$ and its partial derivatives are consequences of their series representation (B.1) with uniformly exponentially bounded coefficient matrices and the finite fourth moment of $Y_{st,k}^{(h)}$ and $\Delta Y_k^{(h)}$ due to Assumption (A3) and Marquardt and Stelzer [38, Proposition 3.30].

(c) A consequence of (B.1) is that both $\partial_i^{st}\varepsilon_k^{(h)}(\vartheta^0)$ and $\partial_{i,j}^{st}\varepsilon_k^{(h)}(\vartheta^0)$ are elements of the Hilbert space generated by $\{Y_l^{(h)}, -\infty<l<k\}$. But $\varepsilon_k^{(h)}(\vartheta^0)$ is orthogonal to the Hilbert space generated by $\{Y_l^{(h)}, -\infty<l<k\}$ so that the

statements follow.
(d) is a conclusion of the construction of the linear innovations by the Kalman filter. □

Appendix C: Proof of Proposition 2.5

First, we present some auxiliary results for the proof of Proposition 2.5.

Lemma C.1. *Let Assumption A and B hold. Define*

$$X_1^{(h)}(\vartheta) = \sum_{j=0}^{\infty} (\mathrm{e}^{A_\vartheta h} - K_\vartheta^{(h)} C_\vartheta)^j K_\vartheta^{(h)} Y_{-j}^{(h)}.$$

Then, the following results are valid.

(a) $\mathbb{E}\left(\sup_{\vartheta\in\Theta} \|X_1^{(h)}(\vartheta)\|^2\right) < \infty$ *and* $\max_{k\in\mathbb{N}}\left\{\frac{1}{(1+k)}\mathbb{E}\left(\sup_{\vartheta\in\Theta}\|\widehat{\varepsilon}_k^{(h)}(\vartheta)\|^2\right)\right\} < \infty.$

(b) $\mathbb{E}\left(\sup_{\vartheta\in\Theta} \|\partial_u X_1^{(h)}(\vartheta)\|^2\right) < \infty$ *and* $\max_{k\in\mathbb{N}}\left\{\frac{1}{(1+k)}\mathbb{E}\left(\sup_{\vartheta\in\Theta}\|\partial_u\widehat{\varepsilon}_k^{(h)}(\vartheta)\|^2\right)\right\} < \infty.$

(c) $\mathbb{E}\left(\sup_{\vartheta\in\Theta} \|\partial_{u,v} X_1^{(h)}(\vartheta)\|^2\right) < \infty$ *and* $\max_{k\in\mathbb{N}}\left\{\frac{1}{(1+k)}\mathbb{E}\left(\sup_{\vartheta\in\Theta}\|\partial_{u,v}\widehat{\varepsilon}_k^{(h)}(\vartheta)\|^2\right)\right\} < \infty.$

Proof. We prove (a) exemplary for (b) and (c). First, remark that

$$\mathbb{E}\|Y_j^{(h)}\|^2 \leq \mathfrak{C}(1+|j|) \quad \text{for } j \in \mathbb{Z}.$$

Since all eigenvalues of $(\mathrm{e}^{A_\vartheta h} - K_\vartheta^{(h)} C_\vartheta)$ lie inside the unit circle (see Scholz [52, Lemma 4.6.7]) and all matrix functions are continuous on the compact set Θ and, hence, bounded, we receive for some $0 < \rho < 1$ that $\sup_{\vartheta\in\Theta}\|\mathrm{e}^{A_\vartheta h} - K_\vartheta^{(h)} C_\vartheta\| \leq \rho$ and $\sup_{\vartheta\in\Theta}\|X_1^{(h)}(\vartheta)\| \leq \mathfrak{C}\sum_{j=0}^\infty \rho^j \|Y_{-j}^{(h)}\|$. Thus,

$$\begin{aligned}
\mathbb{E}\left(\sup_{\vartheta\in\Theta}\|X_1^{(h)}(\vartheta)\|^2\right) &\leq \mathfrak{C}\left(\sum_{j=0}^\infty \rho^j (\mathbb{E}\|Y_{-j}^{(h)}\|^2)^{1/2}\right)^2 \\
&\leq \mathfrak{C}\left(\sum_{j=0}^\infty \rho^j (1+j)^{1/2}\right)^2 < \infty.
\end{aligned}$$

Similarly,

$$\sup_{\vartheta\in\Theta}\|\widehat{\varepsilon}_k^{(h)}(\vartheta)\| \leq \|Y_k\| + \mathfrak{C}\rho^{k-1}\sup_{\vartheta\in\Theta}\|\widehat{X}_1^{(h)}(\vartheta)\| + \sum_{j=1}^{k-1}\mathfrak{C}\rho^j\|Y_{k-j}^{(h)}\|,$$

such that

$$
\begin{aligned}
&\mathbb{E}\left(\sup_{\vartheta\in\Theta}\|\widehat{\varepsilon}_k^{(h)}(\vartheta)\|^2\right)\\
&\leq\quad 3\mathbb{E}\|Y_k\|^2+3\mathfrak{C}^2\rho^{2k-2}\mathbb{E}\left(\sup_{\vartheta\in\Theta}\|\widehat{X}_1^{(h)}(\vartheta)\|^2\right)+3\left(\sum_{j=1}^{k-1}\mathfrak{C}\rho^j(\mathbb{E}\|Y_{k-j}^{(h)}\|^2)^{1/2}\right)^2\\
&\leq\quad \mathfrak{C}\left((1+k)+\rho^{2k-2}\mathbb{E}\left(\sup_{\vartheta\in\Theta}\|\widehat{X}_1^{(h)}(\vartheta)\|^2\right)+k\left(\sum_{j=0}^{\infty}\rho^j\right)^2\right).
\end{aligned}
$$

Finally, due to Assumption B

$$
\begin{aligned}
&\max_{k\in\mathbb{N}}\left\{\frac{1}{(1+k)}\mathbb{E}\left(\sup_{\vartheta\in\Theta}\|\widehat{\varepsilon}_k^{(h)}(\vartheta)\|^2\right)\right\}\\
&\qquad\leq\mathfrak{C}\left(1+\mathbb{E}\left(\sup_{\vartheta\in\Theta}\|\widehat{X}_1^{(h)}(\vartheta)\|^2\right)+\left(\sum_{j=0}^{\infty}\rho^j\right)^2\right)<\infty.
\end{aligned}
$$

□

Lemma C.2. *Let Assumption A and B hold. Furthermore, let* $u,v\in\{1,\dots,s\}$.

(a) Then, there exist a positive random variable ζ *with* $\mathbb{E}(\zeta^2)<\infty$ *and a constant* $0<\rho<1$ *so that* $\sup_{\vartheta\in\Theta}\|\widehat{\varepsilon}_k^{(h)}(\vartheta)-\varepsilon_k^{(h)}(\vartheta)\|\leq\mathfrak{C}\rho^{k-1}\zeta$ *for any* $k\in\mathbb{N}$.

(b) Then, there exist a positive random variable $\zeta^{(u)}$ *with* $\mathbb{E}(\zeta^{(u)})^2<\infty$ *and a constant* $0<\rho<1$ *so that* $\sup_{\vartheta\in\Theta}\|\partial_u\widehat{\varepsilon}_k^{(h)}(\vartheta)-\partial_u\varepsilon_k^{(h)}(\vartheta)\|\leq\mathfrak{C}\rho^{k-1}\zeta^{(u)}$ *for any* $k\in\mathbb{N}$.

(c) Then, there exist a positive random variable $\zeta^{(u,v)}$ *with* $\mathbb{E}(\zeta^{(u,v)})^2<\infty$ *and a constant* $0<\rho<1$ *so that* $\sup_{\vartheta\in\Theta}\|\partial_{u,v}\widehat{\varepsilon}_k^{(h)}(\vartheta)-\partial_{u,v}\varepsilon_k^{(h)}(\vartheta)\|\leq\mathfrak{C}\rho^{k-1}\zeta^{(u,v)}$ *for any* $k\in\mathbb{N}$.

Proof. (a) We use the representation

$$
\widehat{\varepsilon}_k^{(h)}(\vartheta)-\varepsilon_k^{(h)}(\vartheta)=C_\vartheta(\mathrm{e}^{A_\vartheta h}-K_\vartheta^{(h)}C_\vartheta)^{k-1}(\widehat{X}_1^{(h)}(\vartheta)-X_1^{(h)}(\vartheta))
$$

and define $\zeta:=\sup_{\vartheta\in\Theta}\|\widehat{X}_1^{(h)}(\vartheta)\|+\sup_{\vartheta\in\Theta}\|X_1^{(h)}(\vartheta)\|$. Due to Assumption B and Lemma C.1(a) we know that $\mathbb{E}(\zeta^2)<\infty$. Since all eigenvalues of $(\mathrm{e}^{A_\vartheta h}-K_\vartheta^{(h)}C_\vartheta)$ lie inside the unit circle and C_ϑ is bounded as a continuous function on the compact set Θ there exists constants $\mathfrak{C}>0$ and $0<\rho<1$ so that $\sup_{\vartheta\in\Theta}\|C_\vartheta(\mathrm{e}^{A_\vartheta h}-K_\vartheta^{(h)}C_\vartheta)^{k-1}\|\leq\mathfrak{C}\rho^{k-1}$ and

$$
\begin{aligned}
&\sup_{\vartheta\in\Theta}\|\widehat{\varepsilon}_k^{(h)}(\vartheta)-\varepsilon_k^{(h)}(\vartheta)\|\\
&\quad\leq\quad\sup_{\vartheta\in\Theta}\|C_\vartheta(\mathrm{e}^{A_\vartheta h}-K_\vartheta^{(h)}C_\vartheta)\|^{k-1}\sup_{\vartheta\in\Theta}\|\widehat{X}_1^{(h)}(\vartheta)-X_1^{(h)}(\vartheta)\|
\end{aligned}
$$

$$\leq \quad \mathfrak{C}\rho^{k-1}\zeta.$$

(b, c) can be proven similarly. □

Proof of Proposition 2.5.
(a) First,

$$\begin{aligned}\widehat{\mathcal{L}}_n^{(h)}(\vartheta) - \mathcal{L}_n^{(h)}(\vartheta) \quad &= \quad \frac{1}{n}\sum_{k=1}^{n}\Big[(\widehat{\varepsilon}_k^{(h)}(\vartheta) - \varepsilon_k^{(h)}(\vartheta))^\mathsf{T}\big(V_\vartheta^{(h)}\big)^{-1}\widehat{\varepsilon}_k^{(h)}(\vartheta) \\ &\qquad -\varepsilon_k^{(h)}(\vartheta)^\mathsf{T}\big(V_\vartheta^{(h)}\big)^{-1}(\varepsilon_k^{(h)}(\vartheta) - \widehat{\varepsilon}_k^{(h)}(\vartheta))\Big].\end{aligned}$$

Then,

$$\begin{aligned}&n\sup_{\vartheta\in\Theta}|\widehat{\mathcal{L}}_n^{(h)}(\vartheta) - \mathcal{L}_n^{(h)}(\vartheta)| \\ &\leq \sup_{\vartheta\in\Theta}\|(V_\vartheta^{(h)})^{-1}\|\sum_{k=1}^{n}\sup_{\vartheta\in\Theta}\|\widehat{\varepsilon}_k^{(h)}(\vartheta) - \varepsilon_k^{(h)}(\vartheta)\|\left(\sup_{\vartheta\in\Theta}\|\widehat{\varepsilon}_k^{(h)}(\vartheta)\| + \sup_{\vartheta\in\Theta}\|\varepsilon_k^{(h)}(\vartheta)\|\right).\end{aligned}$$

Due to Lemma B.1 and Lemma C.2(a)

$$\leq \mathfrak{C}\zeta\sum_{k=1}^{n}\rho^k\left[\sup_{\vartheta\in\Theta}\|\widehat{\varepsilon}_k^{(h)}(\vartheta)\| + \sup_{\vartheta\in\Theta}\|\varepsilon_k^{(h)}(\vartheta)\|\right]$$

with $\mathbb{E}(\zeta^2) < \infty$. From this and Cauchy-Schwarz inequality we can conclude that

$$\begin{aligned}&n\mathbb{E}\left(\sup_{\vartheta\in\Theta}|\widehat{\mathcal{L}}_n^{(h)}(\vartheta) - \mathcal{L}_n^{(h)}(\vartheta)|\right) \\ &\qquad \leq \mathfrak{C}(\mathbb{E}\zeta^2)^{1/2}\sum_{k=1}^{n}\rho^k\left[\mathbb{E}\left(\sup_{\vartheta\in\Theta}\|\widehat{\varepsilon}_k^{(h)}(\vartheta)\|^2\right)^{1/2} + \mathbb{E}\left(\sup_{\vartheta\in\Theta}\|\varepsilon_k^{(h)}(\vartheta)\|^2\right)^{1/2}\right].\end{aligned}$$

An application of Lemma C.1(a) yields

$$\leq \mathfrak{C}(\mathbb{E}\zeta^2)^{1/2}\sum_{k=1}^{\infty}\rho^k(1+k)^{1/2} < \infty.$$

This proves, $n\sup_{\vartheta\in\Theta}|\widehat{\mathcal{L}}_n^{(h)}(\vartheta) - \mathcal{L}_n^{(h)}(\vartheta)| = \mathcal{O}_p(1)$ so that (a) follows. Again (b) and (c) can be proven similarly. □

Acknowledgement

We would like to thank Eckhard Schlemm and Robert Stelzer, who kindly provided the MATLAB code for the simulation and parameter estimation of stationary state space processes. We adapted the code in order to include not only stationary state space models, but also cointegrated state space models.

References

[1] Andresen, A., Benth, F. E., Koekebakker, S. and Zakamulin, V. (2014). The CARMA interest rate model. *Int. J. Theor. Appl. Finance* **17** 1–27. MR3198712

[2] Aoki, M. (1990). *State Space Modeling of Time Series*, 2nd ed. Springer, Berlin. MR1073171

[3] Barndorff-Nielsen, O. E. (1997). Normal inverse Gaussian distributions and stochastic volatility modelling. *Scand. J. Stat.* **24** 1–13. MR1436619

[4] Bauer, D. and Wagner, M. (2002). Asymptotic Properties of Pseudo Maximum Likelihood Estimates for Multiple Frequency I(1) Processes Diskussionsschriften, Universitaet Bern, Departement Volkswirtschaft.

[5] Bauer, D. and Wagner, M. (2002). Estimating cointegrated systems using subspace algorithms. *J. Econometrics* **111** 47–84. MR1926537

[6] Benth, F. E., Klüppelberg, C., Müller, G. and Vos, L. (2014). Futures pricing in electricity markets based on stable CARMA spot models. *Energ. Econ.* **44** 392–406.

[7] Bergstrom, A. R. (1990). *Continuous Time Econometric Modelling*. Oxford University Press, Oxford.

[8] Bergstrom, A. R. (1997). Gaussian estimation of mixed-order continuous-time dynamic models with unobservable stochastic trends from mixed stock and flow data. *Econometric Theory* **13** 467–505. MR1475068

[9] Bernstein, D. S. (2009). *Matrix Mathematics: Theory, Facts, and Formulas*, 2nd ed. Princeton Univ. Press, Princeton. MR2513751

[10] Blevins, J. R. (2017). Identifying restrictions for finite parameter continuous time models with discrete time data. *Econometric Theory* **33** 739–754. MR3637974

[11] Boubacar Maïnassara, Y. and Francq, C. (2011). Estimating structural VARMA models with uncorrelated but non-independent error terms. *J. Multivar. Anal.* **102** 496–505. MR2755011

[12] Bradley, R. C. (2007). *Introduction to Strong Mixing Conditions* **1**. Kendrick, Heber City. MR2325295

[13] Brockwell, P. J. and Davis, R. A. (1998). *Time Series: Theory and Methods*, 2nd ed. *Springer Ser. Statist.* Springer, New York. MR2839251

[14] Chambers, A. J. and McCrorie, J. R. (2007). Frequency domain estimation of temporally aggregated Gaussian cointegrated systems. *J. Econometrics* **136** 1–29. MR2328583

[15] Chambers, M. J., McCrorie, J. R. and Thornton, M. A. (2018). Continuous time modelling based on an exact discrete time representation. In *Continuous Time Modelling in the Behavioral and Related Sciences* (K. van Montford, J. Oud and M. Voelkle, eds.) Springer, New York. 317–357.

[16] Chambers, M. J. and Thornton, M. A. (2012). Discrete time representation of continuous time ARMA processes. *Econometric Theory* **28** 219–238. MR2899219

[17] Chen, F., Agüero, J. C., Gilson, M., Garnier, H. and Liu, T. (2017).

EM-based identification of continuous-time ARMA Models from irregularly sampled data. *Automatica* **77** 293–301. MR3605789
[18] Fasen-Hartmann, V. and Scholz, M. (2019). Cointegrated continuous-time linear state-space and MCARMA models. *Stochastics* to appear.
[19] Fasen-Hartmann, V. and Scholz, M. (2019). A canonical form for a cointegrated MCARMA process. In preparation.
[20] Fasen, V. (2013). Time series regresion on integrated continuous-time processeswith heavy and light tails. *Econometric Theory* **29** 28–67. MR3028879
[21] Fasen, V. (2014). Limit theory for high frequency sampled MCARMA models. *Adv. Appl. Probab.* **46** 846–877. MR3254345
[22] Garnier, H. and Wang, L., eds. (2008). *Identification of Continuous-time Models from Sampled Data*. *Advances in Industrial Control*. Springer, London.
[23] Guidorzi, R. (1975). Canonical structures in the identification of multivariable systems. *Automatica* **11** 361–374. MR0444226
[24] Hannan, E. J. and Deistler, M. (2012). *The Statistical Theory of Linear Systems*. Society for Industrial and Applied Mathematics, Philadelphia. MR3397291
[25] Hansen, L. P. and Sargent, T. J. (1983). The dimensionality of the aliasing problem in models with rational spectral densities. *Econometrica* **51** 377–387. MR0688733
[26] Harvey, A. C. and Stock, J. H. (1985). The estimation of higher-order continuous time autoregressive models. *Econometric Theory* **1** 97–112.
[27] Harvey, A. C. and Stock, J. H. (1985). Continuous time autoregressive models with common stochastic trends. *J. Econom. Dynam. Control.* **12** 365–384. MR0986521
[28] Harvey, A. C. and Stock, J. H. (1989). Estimating integrated higher-order continuous time autoregressions with an application to money-income causality. *J. Econometrics* **42** 319–336. MR1040747
[29] Ibragimov, I. A. (1962). Some limit theorems for stationary processes. *Theory Probab. Appl.* **7** 349–382. MR0148125
[30] Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* **59** 1551–1580. MR1135641
[31] Johansen, S. (1995). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. *Advanced Texts in Econometrics*. Oxford Univ. Press, Oxford. MR1487375
[32] Kessler, M. and Rahbek, A. (2001). Asymptotic likelihood based inference for co-integrated homogenous Gaussian diffusions. *Scand. J. Statist.* **28** 455–470. MR1858411
[33] Kessler, M. and Rahbek, A. (2004). Identification and inference for multivariate cointegrated and ergodic Gaussian diffusions. *Stat. Inference Stoch. Process.* **7** 137–151. MR2061182
[34] Krengel, U. (1985). *Ergodic Theorems: With a Supplement*. De Gruyter, Berlin. MR0797411

[35] Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis.* Springer, Berlin. MR2172368

[36] Lütkepohl, H. and Claessen, H. (1997). Analysis of cointegrated VARMA processes. *J. Econometrics* **80** 223–239. MR1483178

[37] Lütkepohl, H. and Poskitt, D. S. (1996). Specification of echelon-form VARMA models. *J. Bus. Econom. Statist.* **14** 69–79. MR1433592

[38] Marquardt, T. and Stelzer, R. (2007). Multivariate CARMA processes. *Stochastic Process. Appl.* **117** 96–120. MR2287105

[39] Masuda, H. (2004). On multidimensional Ornstein-Uhlenbeck processes driven by a general Lévy process. *Bernoulli* **10** 97–120. MR2044595

[40] McCrorie, J. R. (2003). The problem of aliasing in identifying finite parameter continuous time stochastic models. *Acta Applicandae Mathematicae* **79** 9–16. MR2021872

[41] McCrorie, J. R. (2009). Estimating continuous-time models on the basis of discrete data via an exact discrete analog. *Econometric Theory* **25** 1120–1137. MR2515474

[42] Phillips, P. C. B. (1973). The problem of identification in finite parameter continuous time models. *J. Econometrics* **1** 351–362. MR0452554

[43] Phillips, P. C. B. (1991). Error correction and long-run equilibrium in continuous time. *Econometrica* **59** 967–980. MR1113542

[44] Reinsel, G. C. (1997). *Elements of Multivariate Time Series Analysis*, 2nd ed. Springer-Verlag, New York. MR1451875

[45] Saikkonen, P. (1992). Estimation and testing of cointegrated systems by an autoregressive approximation. *Econometric Theory* **8** 1–27. MR1162612

[46] Saikkonen, P. (1993). Continuous weak convergence and stochastic equicontinuity results for integrated processes with an application to the estimation of a regression model. *Econometric Theory* **9** 155–188. MR1229412

[47] Saikkonen, P. (1995). Problems with the asymptotic theory of maximum likelihood estimation in integrated and cointegrated systems. *Econometric Theory* **11** 888–911. MR1458944

[48] Sato, K.-I. (1999). *Lévy Processes and Infinitely Divisible Distributions. Cambridge Stud. Adv. Math.* **68**. Cambridge Univ. Press, Cambridge. MR1739520

[49] Schlemm, E. (2011). Estimation of Continuous-Time ARMA Models and Random Matrices with Dependent Entries, Dissertation, Technische Universität München.

[50] Schlemm, E. and Stelzer, R. (2012). Multivariate CARMA processes, continuous-time state space models and complete regularity of the innovations of the sampled processes. *Bernoulli* **18** 46–63. MR2888698

[51] Schlemm, E. and Stelzer, R. (2012). Quasi maximum likelihood estimation for strongly mixing state space models and multivariate Lévy-driven CARMA processes. *Electron. J. Stat.* **6** 2185–2234. MR3020261

[52] Scholz, M. (2016). Estimation of Cointegrated Multivariate Continuous-time Autoregressive Moving Average Processes, Dissertation, Karlsruher Institute of Technology (KIT), Karlsruhe.

[53] Sinha, N. K. and Rao, G. P. (1991). *Identification of Continuous-Time Systems: Methodology and Computer Implementation*. Springer, Dordrecht.

[54] Sinha, N. K. and Rao, G. P. (1991). Continuous-time models and approaches. In *Identification of Continuous-Time Systems: Methodology and Computer Implementation* (N. K. Sinha and G. P. Rao, eds.) 1–15. Springer, Dordrecht.

[55] Stockmarr, A. and Jacobsen, M. (1994). Gaussian diffusions and autoregressive processes: weak convergence and statistical inference. *Scand. J. Statist.* **21** 403–429. MR1310085

[56] Thornton, M. A. and Chambers, M. J. (2016). The exact discretisation of CARMA models with applications in finance. *J. Emp. Finance* **38** 739–761.

[57] Thornton, M. A. and Chambers, M. J. (2017). Continuous time ARMA processes: discrete time representation and likelihood evaluation. *J. Econom. Dynam. Control* **79** 48–65. MR3649332

[58] Todorov, V. (2009). Estimation of continuous-time stochastic volatility models with jumps using high-frequency data. *J. Econometrics* **148** 131–148. MR2500652

[59] Yap, S. F. and Reinsel, G. C. (1995). Estimation and testing for unit roots in a partially nonstationary vector autoregressive moving average model. *J. Amer. Statist. Assoc.* **90** 253–267. MR1325133

[60] Zadrozny, P. (1988). Gaussian likelihood of continuous-time ARMAX models when data are stocks and flows at different frequencies. *Econometric Theory* **4** 108–124.