Asymptotic Properties of Pseudo Maximum Likelihood Estimates for Multiple Frequency I(1) Processes

Dietmar Bauer
Martin Wagner

02-05

June 2002

Diskussionsschriften

Universität Bern
Volkswirtschaftliches Institut
Gesellschaftstrasse 49
3012 Bern, Switzerland
Tel: 41 (0)31 631 45 06
Web: www-vwi.unibe.ch

Asymptotic Properties of Pseudo Maximum Likelihood Estimates for Multiple Frequency I(1) Processes

Dietmar Bauer *
Inst. for Econometrics,
Operations Research and System Theory
TU Wien
Argentinierstr. 8, A-1040 Wien

Martin Wagner
Department of Economics
University of Bern
Gesellschaftsstrasse 49
CH-3012 Bern

Abstract

In this paper we derive (weak) consistency and the asymptotic distribution of pseudo maximum likelihood estimates for multiple frequency I(1) processes. By multiple frequency I(1) processes we denote processes with unit roots at arbitrary points on the unit circle with the integration orders corresponding to these unit roots all equal to 1. The parameters corresponding to the cointegrating spaces at the different unit roots are estimated super-consistently and have a mixture of Brownian motions limiting distribution. All other parameters are asymptotically normally distributed and are estimated at the standard square root of T rate.
The problem is formulated in the state space framework, using the canonical form and parameterization introduced by Bauer and Wagner (2002b). Therefore the analysis covers vector ARMA processes and is not restricted to autoregressive processes.

JEL Classification: C13, C32
Keywords: State space representation, unit roots, cointegration, pseudo maximum likelihood estimation

*Support by the Austrian FWF under the project number P-14438-INF is gratefully acknowledged. Corresponding author, e-mail:Dietmar.Bauer@tuwien.ac.at, fax:++43 +1 58801 11944.

1 Introduction

During the last decades the modelling of trends and seasonal components with time series models that allow for unit roots and seasonal unit roots has become quite popular. Cointegration and seasonal cointegration have become prominent tools, formalizing the observation that both trends and persistent seasonal fluctuations that cannot be appropriately modelled with deterministic components may be present in the data.

Usually the unit roots literature is formulated in terms of (vector) autoregressive or (vector) autoregressive moving average (ARMA) models. The state space framework, which is – in a specific sense discussed below – equivalent to the ARMA framework has not obtained a lot of attention. Early exceptions are given by Aoki (1990) and Aoki and Havenner (1997), who however focus on the only unit root $z = 1$. In a recent paper, Bauer and Wagner (2002b), it has been shown that the state space framework can be used to obtain a convenient representation as well as a parameterization of rational unit root processes with integer integration orders corresponding to the different unit roots. The results derived in that paper form the basis for the statistical analysis presented below. The advantage of the state space representation is, as will be seen below, that it directly leads to system representations where the contributions corresponding to the various unit roots are separated in a Granger type representation. Note that similar results that separate the nonstationary components of different integration orders and corresponding to different unit roots are also directly obtained for the general case of processes with higher integration orders, see Bauer and Wagner (2002b).

Estimation results are presented in this paper for the class of processes where the integration orders (see Section 2 for precise definitions) are equal to 1 for all unit roots, which we call multiple frequency I(1), or short MFI(1) processes. The results are based upon optimizing the Gaussian (pseudo) likelihood over the parameter set that is given from the developed parameterization. The first result is the consistency proof for the pseudo likelihood estimate. It is then shown that the parameters corresponding to the cointegrating spaces are estimated super-consistently and all other parameters are estimated with the standard $\sqrt{T}$ rate, T denoting the sample size. Furthermore also the asymptotic distribution is derived. The parameters corresponding to the cointegrating spaces have an asymptotic distribution consisting of a mixture of Brownian motions and all other parameters are asymptotically normally distributed.

Previous results differ in two aspects from the results presented below: A part of the literature is focusing only on the unit root $z = 1$, this includes e.g. Yap and Reinsel (1995), who derive the maximum likelihood estimator for Gaussian I(1) ARMA processes, or the work of Johansen (1995) for AR processes. Also the (efficient) regression based approaches of Phillips (1991a, 1991b) deal with processes integrated only at $z = 1$. The part of the literature dealing with estimation for processes with arbitrary (but known) locations of the unit roots of integration orders equal to 1, is developed for AR processes and not for ARMA processes, see e.g. Lee (1992), Johansen and Schaumburg (1999) or Gregoir (1999b).

Our results, or more specifically the way they are proven, are inspired by results derived in Saikkonen (1993, 1995). In the consistency proof we extend (in Lemma 3) his stochastic equicontinuity results from I(1) processes to MFI(1) processes. In the derivation of the asymptotic distribution we draw from Chan and Wei (1988) and the algebraically more convenient complex valued version of their results in Johansen and Schaumburg (1999).

A limitation of the results of the paper is the fact that it is assumed that the locations as well as the integration orders of the unit roots are known or correctly specified. Hence, tests for these integer valued parameters are called for and likelihood ratio tests would be prime candidates. However, the material in the paper relies heavily on the assumption of a correct specification of the unit root structure (for a definition see below), which makes the extension to misspecification analysis at least not straightforward. There exist various options to overcome that limitation. For the I(1) case many alternatives are known, both for the AR as well as the ARMA case. One alternative, related to the present paper as it is also formulated in the state space framework, is presented in Bauer and Wagner (2002a). In that paper so called subspace algorithms are used to estimate I(1) processes and to test for the dimension of the cointegrating space. The computationally cheap subspace algorithms can also be used as consistent initial estimates for pseudo maximum likelihood estimation. Subspace algorithms are not yet analyzed for processes with unit roots other than $z = 1$. For a general configuration of unit roots with corresponding integration orders all equal to 1, Johansen and Schaumburg (1999) derive tests for determining the dimensions of the cointegrating spaces for AR processes. It appears possible to extend their result to the ARMA case; compare e.g. the extension of Johansen's AR approach for $z = 1$ to ARMA models by Saikkonen (1992).

The canonical form and parameterization developed in Bauer and Wagner (2002b) covers more general cases than analyzed in this paper. The statistical analysis of systems with higher integration orders, potentially including tests for the cointegrating ranks is an important topic of further research.

The paper is organized as follows: In the following section we present the model class and review some basic facts about state space representations. In Section 3 the canonical form is presented for the special class of processes we are dealing with in this paper. In this section also a small example illustrating the canonical form and the parameterization is discussed. In Section 4 the estimates and their asymptotic properties are discussed. Section 5 summarizes and concludes with a brief discussion of the results. The proofs are deferred to Appendix A. In Appendix B for completeness' sake one parameterization for complex positive lower triangular matrices, which are used throughout the paper, is presented.

Notation in the paper is as follows: $\mathbb{N}$ denotes the integers, $\mathbb{R}$ the real numbers, $\mathbb{C}$ the complex numbers, $\mathbb{P}$ denotes probability and $\mathbb{E}$ expectation. For both, vectors and matrices $x \in \mathbb{C}^{m\times n}$ the complex conjugate transpose is denoted by x' and the complex conjugate is denoted by $\bar{x}$. I_d denotes the d-dimensional identity matrix and $0^{m\times n}$ the m times n zero matrix. $\rightarrow$ denotes convergence in probability and $\stackrel{d}{\rightarrow}$ denotes convergence in distribution. For a sequence of random variables f_T and a sequence of constants g_T, the expression $f_T = o_P(g_T)$ means $f_T/g_T \rightarrow 0$ in probability. $\|.\|$ denotes the two norm both for vectors and for matrices, both real and complex. Finally i denotes $\sqrt{-1}$, unless explicitly stated differently.

2 The State Space Framework

This section discusses the model set and the assumptions and defines the class of processes under study. Some links between state space and ARMA representations are briefly discussed. Readers unacquainted with the state space framework are referred to Hannan and Deistler (1988, Chapters 1 and 2) for a precise discussion of the links for the stationary case. We consider in this paper finite dimensional, time invariant, discrete time systems in their state space representation of the form:

$$\begin{array}{rcl} y_t & = & Cx_t + Ds_t + \varepsilon_t \\ x_{t+1} & = & Ax_t + B\varepsilon_t \end{array} \tag{1}$$

where $y_t \in \mathbb{R}^s, t \in \mathbb{N}$ denotes the s-dimensional output, observed for $t = 1, \ldots, T$. $x_t \in \mathbb{C}^n$ denotes the n-dimensional unobserved state vector and $s_t \in \mathbb{C}^m$ accounts for deterministic variables. The deterministic variables contain the constant as well as seasonal cycles to some frequencies $\omega_j \in [0, 2\pi)$. For the corresponding component of the variable s_t, say $s_{t,j}$, then

$s_{t+1,j} = z_j s_{t,j}$ holds, with $z_j = e^{i\omega_j}$ and $s_{1,j} = 1$. The constant corresponds, using this notation, to the frequency 0. The initial state x_1 is assumed to be constant. Concerning $\varepsilon_t, t \in \mathbb{N}$ we assume that it is a strictly stationary ergodic white noise sequence for which the following conditions hold.

$$\mathbb{E}\{\varepsilon_t | \mathcal{F}_{t-1}\} = 0 \tag{2}$$

$$\mathbb{E}\{\varepsilon_t \varepsilon_t' | \mathcal{F}_{t-1}\} = \mathbb{E}\{\varepsilon_t \varepsilon_t'\} = \Sigma^0 \tag{3}$$

$$\mathbb{E}\varepsilon_{t,a}^4 < \infty \tag{4}$$

where $\varepsilon_{t,a}$ denotes the a-th component of the vector ε_t and $\mathcal{F}_{t-1}$ denotes the σ-algebra spanned by the past, i.e. by $\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots, \varepsilon_0$. The above conditions on ε_t are referred to as *standard conditions* throughout the paper.

$A \in \mathbb{C}^{n \times n}, B \in \mathbb{C}^{n \times s}, C \in \mathbb{C}^{s \times n}, D \in \mathbb{C}^{s \times m}$ are complex matrices. Usually it is assumed that x_t, A, B, C and D are real rather than complex, however for (some of) the following results the use of complex quantities simplifies the algebra considerably. Real valuedness of the output imposes a number of restrictions on the system matrices (A, B, C, D), see below for a discussion of the relations between real valued and complex valued system descriptions.

It is straightforward to verify that for given initial value x_1 the solution to the system of VDEs (1) is given by:

$$y_t = Cx_t + Ds_t + \varepsilon_t = CAx_{t-1} + Ds_t + \varepsilon_t + CB\varepsilon_{t-1} = \ldots = CA^{t-1}x_1 + Ds_t + \sum_{j=0}^{t-1} K_j \varepsilon_{t-j} \tag{5}$$

The matrix sequence $K_j = CA^{j-1}B, j \geq 1, K_0 = I$ denotes the so called sequence of impulse response coefficients. Let $\lambda_{max}(A)$ denote an eigenvalue of A of maximum modulus. Then for $z \in \mathbb{C}, |z| < |\lambda_{max}(A)|^{-1} - \epsilon, \epsilon > 0$, the transfer function $k(z) = \sum_{j=0}^{\infty} K_j z^j$ converges absolutely and has the representation $k(z) = I + zC(I - zA)^{-1}B$. The matrix triple (A, B, C) is called a *state space realization* of the transfer function $k(z)$ just defined. It follows by construction that the resulting $k(z)$ is a rational function. Let $\tilde{S}_n$ denote the set of all matrix triples (A, B, C), which correspond to state dimension n, i.e. $A \in \mathbb{C}^{n \times n}$ and with B and C of respective dimensions. Define the mapping $\Pi(A, B, C) \to k(z) = I + zC(I - zA)^{-1}B$ linking the matrix triple to the corresponding transfer function.

Also conversely, for each rational function $k(z), k(0) = I$, there exists a state space realization, i.e. a matrix triple (A, B, C) such that $k(z) = \Pi(A, B, C)$, see Hannan and Deistler (1988, Chapter 1). It is also a well known fact that for every rational transfer function with $k(0) = I$ also representations as an ARMA system exist, i.e. there exist matrix polynomials $a(z) = \sum_{j=0}^{p} A_j z^j, A_0 = I, A_p \neq 0$, $b(z) = \sum_{j=0}^{q} B_j z^j, B_0 = I, B_q \neq 0$ such that $k(z) = a^{-1}(z)b(z)$ holds. We denote the corresponding mapping attaching the transfer function $k(z) = a^{-1}(z)b(z)$ to the matrix polynomials $(a(z), b(z))$ with $\bar{\Pi}$. Hence, for every ARMA system $(a(z), b(z))$ there exists a state space system (A, B, C), such that $\Pi(A, B, C) = \bar{\Pi}(a(z), b(z))$. Both, state space as well as ARMA representations of a transfer function $k(z)$ are not unique. For a fixed transfer function $k(z)$ the sets $\{(A, B, C) \in \bigcup_{j=0}^{\infty} \tilde{S}_j : \Pi(A, B, C) = k(z)\}$ and $\{(a, b) : \bar{\Pi}(a, b) = k(z)\}$ are called *equivalence sets*. For state space systems *observationally equivalent* representations are obtained by transforming any representation with nonsingular matrices T, as $\Pi(A, B, C) = \Pi(TAT^{-1}, TB, CT^{-1})$. For ARMA systems polynomial matrices have the same function, since $\bar{\Pi}(pa, pb) = \bar{\Pi}(a, b)$ for all polynomial matrices $p(z)$ with $p(0) = I$.

A state space realization of a transfer function $k(z)$ is called *minimal*, if no observationally equivalent state space realization with smaller state dimension exists. The concept of minimality is linked to three matrices: The *observability* matrix $\mathcal{O} = [C', A'C', (A^2)'C', \ldots]'$, the *controllability* matrix $\mathcal{C} = [B, AB, A^2B, \ldots]$ and the *Hankel* matrix $\mathcal{H} = \mathcal{O}\mathcal{C} = [CA^{i+j-2}B]_{i,j=1,\ldots}$. For given $(A, B, C) \in \tilde{S}_n$ it is easy to see that the rank of $\mathcal{H}$ is at most n. Hence the rank of $\mathcal{H}$ is a lower bound for the minimal state dimension. It can be shown that there always exists a realization achieving this minimal dimension, hence the minimal order, i.e. the minimal state dimension, is equal to the rank of $\mathcal{H}$. This integer n is equal to the McMillan degree of the transfer function (see e.g. Hannan and Deistler, 1988). Under the assumption of minimality, two state space systems (A_1, B_1, C_1) and (A_2, B_2, C_2) are observationally equivalent, if and only if there exists a nonsingular matrix $T \in \mathbb{C}^{n \times n}$ such that $A_1 = TA_2T^{-1}, B_1 = TB_2$ and $C_1 = C_2T^{-1}$. The nonsingular matrices T that generate observationally equivalent state space systems correspond to changes in the state space basis and result in different factorizations of the Hankel matrix, then given by $\mathcal{H} = [\mathcal{O}T^{-1}][T\mathcal{C}]$. The concept corresponding to minimality in the ARMA framework is left coprimeness (see Hannan and Deistler, 1988, Section 2.2).

Thus, up to now we have established that for each rational transfer function $k(z)$ with McMillan degree n there exists a left coprime ARMA representation $(a(z), b(z)), \bar{\Pi}(a(z), b(z)) = k(z)$ and a minimal state space realization $(A, B, C), \Pi(A, B, C) = k(z)$. From the ARMA framework it is well understood that in a left coprime representation the locations of the roots of the determinant of the matrix polynomial $a(z)$ determine the integration or stationarity properties of the resulting ARMA processes. The analogue for minimal state space realizations are the locations of the eigenvalues of A: If the poles of $k(z)$ are defined as the roots of $\det a(z)$ from any left coprime matrix fraction description $(a(z), b(z))$, then λ is a pole of $k(z)$ if and only if $\det(I - \lambda A) = 0$ for any minimal state space realization (A, B, C) of $k(z)$. Hence, if $\lambda \neq 0$ is a pole, then λ^{-1} is an eigenvalue of A. Similarly, if the zeros of the transfer function are defined as the zeros of $\det b(z)$, then λ is a zero of $k(z)$, if and only if $\det(I - \lambda(A - BC)) = 0$. The paper deals only with processes with eigenvalues of A smaller or equal than one in absolute value, this restriction of $|\lambda_{max}(A)| \leq 1$ is called *non-explosiveness* restriction. In terms of an ARMA representation we thus assume $\det(a(z)) \neq 0, |z| < 1$. Similarly we restrict attention to strictly minimum-phase systems, i.e. to systems where $|\lambda_{max}(A - BC)| < 1$ or equivalently to $k(z)$ such that the zeros of $k(z)$ lie outside the closed unit disc. Hence we exclude systems with zeros on the unit circle, which occur e.g. if the time series is overdifferenced. Let us denote the set of all rational transfer functions of McMillan degree n, where the poles are on or outside the closed unit disc and the zeros are outside the open unit disc by M_n.

For minimal state space representation and left coprime ARMA systems not only the transfer function constitutes a link, but also the solutions to the corresponding vector difference equations (VDEs) are closely related (cf. Lemma 1 in Bauer and Wagner, 2002b). It can be shown that for each solution $y_t, t \in \mathbb{N}$ of the system equations (1), there exist initial conditions $y_0, \ldots, y_{-\max\{p,q\}}, \varepsilon_0, \ldots, \varepsilon_{-\max\{p,q\}}$ and deterministic terms Ds_t, such that $y_t - Ds_t$ is a solution to the ARMA equations $a(z)y_t = b(z)\varepsilon_t$. Conversely also for each solution $z_t, t \in \mathbb{N}$ of the ARMA equations, there exist state space realizations (A, B, C), not necessarily of order n, and an initial state x_1 depending on the initial conditions for the ARMA system, such that z_t is a solution to the state space equations. Then defining the deterministic terms suitably, it can be shown that a state space system of order n can be chosen. In this sense state space systems and ARMA systems generate identical processes as solutions.

We are now left to give our definition of integrated processes on $\mathbb{N}$. For this we require the notion of an asymptotically stationary process: A process $u_t, t \in \mathbb{N}$ is called asymptotically stationary, if $u_t = \sum_{j=0}^{t-1} c_j \varepsilon_{t-j}$ for some white noise process $\varepsilon_t, t \in \mathbb{Z}$, such that $v_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j}$ defines a stationary process. We always assume here that $\sum_{j=0}^{\infty} \|c_j\| < \infty$. This is sufficient for v_t to be stationary. For convenience we will furthermore use the sloppy notation $u_t = c(z)\varepsilon_t = \sum_{j=0}^{t-1} c_j \varepsilon_{t-j}$. The difference operator at frequency ω is defined as follows:

$$\Delta_\omega(z) = \begin{cases} 1 - e^{i\omega} z, & \omega \in \{0, \pi\} \\ (1 - e^{i\omega} z)(1 - e^{-i\omega} z), & \omega \in (0, \pi) \end{cases} \tag{6}$$

Further define a linearly deterministic process s_t to be a process such that $\sup_{t>t_0} \mathbb{E}\|s_{t|t_0} - s_t\| = 0$ for some $t_0 \in \mathbb{N}$ where $s_{t|t_0}$ denotes the best linear prediction of s_t, based on $s_1, \dots, s_{t_0}$. In particular all solutions to the homogenous equation $a(z)s_t = 0$ fall into this category. Then integration is defined as follows:

Definition 1 *The s-dimensional real random process y_t has integration or unit root structure $((\omega_1, h_1), \dots, (\omega_l, h_l))$, with $\omega_k \in [0, \pi], h_k > 0$ for $k = 1, \dots, l$, if there exists a linearly deterministic term s_t and a matrix $D \in \mathbb{C}^{s \times m}$ such that*

$$\Delta_{\omega_1}^{h_1}(z) \dots \Delta_{\omega_l}^{h_l}(z) \left[y_t - D s_t \right] = c(z) \varepsilon_t \tag{7}$$

for $c(z)\varepsilon_t = \sum_{j=0}^{t-1} c_j \varepsilon_{t-j}$ corresponding to the Wold representation of the stationary process $v_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j}$ with $c(z) = \sum_{j=0}^{\infty} c_j z^j \neq 0$ for all $|z| = 1$.
If $c(z)$ is a rational function of z, y_t is called a rational process*.*
If the unit root structure is given by $((\omega_1, 1), \dots, (\omega_l, 1))$, the process is called multiple frequency I(1) *or short* MFI(1) *process.*

Remark 1 *In the above definition the integration orders are defined for the s-dimensional vector y_t, not for the individual components. From the requirement $c(e^{i\omega_k}) \neq 0$ for $k = 1, \dots, l$ it follows that for each unit root at least one component is integrated of order h_k. Hence, the integration order is defined as the maximal integration order over all components. In particular in the definition it is not required that all components have the same integration order.*

Remark 2 *The definition given above incorporates the real valuedness of y_t due to the definition of the filters Δ_ω. If one prefers, one can analyze the complex unit roots that occur in pairs of conjugate complex roots separately, by using the filters $(1 - z_k z)^{h_k}$ with $z_k = e^{i\omega_k}$, $\omega_k \in [0, 2\pi)$ for $k = 1, \dots, l_{2\pi}$ say. For later use we partition the deterministic variables $s_t = \left[(s_t^1)', (s_t^2)' \right]$ and collect in $s_t^1 \in \mathbb{C}^{l_{2\pi}}$ the coordinates corresponding the unit roots ω_k, such that $(1 - z_k z) s_{t,k}^1 = 0$ holds for $k = 1, \dots, l_{2\pi}$. In s_t^2 the cyclical components to the non-unit root frequencies are collected.*

Remark 3 *In the definition the linearly deterministic components are subtracted before applying the filters Δ_ω. This implies e.g. that so called trend stationary processes are according to our definition not integrated. Note also that as a consequence of the definition the first difference of a process with integration structure* $(0, 1)$ *is not necessarily asymptotically stationary. Instead this differences will in general have a Wold decomposition with a nonzero linearly deterministic process $D(s_t - s_{t-1})$.*

Let us next define also cointegration, where we present the definition of static and dynamic cointegration for the general case. In this paper we are only interested in MFI(1) processes, hence the integers h_k and h_k^r in the following definition can be either 1 or 0.

Definition 2 *A real valued process y_t with integration structure $((\omega_1, h_1), \dots, (\omega_l, h_l))$ is called* cointegrated or statically cointegrated of order $((\omega_1, h_1, h_1^r), \dots, (\omega_l, h_l, h_l^r))\,, 0 \leq h_k^r \leq h_k, k = 1, \dots, l$, *where* $\max_{k=1,\dots,l}(h_k - h_k^r) > 0$, *if there exists a vector $\beta \in \mathbb{R}^s, \beta \neq 0$, such that $\beta' y_t$ is integrated of order $((\omega_1, h_1^r), \dots, (\omega_l, h_l^r))$.*
A real valued process y_t with integration structure $((\omega_1, h_1), \dots, (\omega_l, h_l))$ is called dynamically cointegrated of order $((\omega_1, h_1, h_1^r), \dots, (\omega_l, h_l, h_l^r))\,, 0 \leq h_k^r \leq h_k, k = 1, \dots, l$, *if there exist vectors $\beta_0, \beta_1 \in \mathbb{R}^s, \beta_0 \neq 0, \beta_1 \neq 0$, such that $\beta_0' y_t + \beta_1' y_{t-1}$ is integrated of order $((\omega_1, h_1^r), \dots, (\omega_l, h_l^r))$, where* $\max_{k=1,\dots,l} \|\beta_0 + \beta_1 z_k\|(h_k - h_k^r) > 0$.

We have already established the fact that the integration properties of the solutions of the state space equations depend upon the eigenvalues of A, this follows also directly from (5) and $K_j = CA^{j-1}B$. Let $J = TAT^{-1}$ denote the Jordan normal form of A (see e.g. Meyer, 2000) for suitable T. In the Jordan normal form, the eigenvalues are directly seen appearing along the diagonal, ordered in Jordan segments corresponding to the different eigenvalues, and the Jordan segments are grouped in Jordan blocks corresponding to chains of generalized eigenvectors. The matrix J has the following structure:

$$J = \begin{bmatrix} J_u & 0 \\ 0 & J_{st} \end{bmatrix} = \begin{bmatrix} J_1 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & J_{l_{2\pi}} & 0 \\ 0 & \dots & 0 & J_{st} \end{bmatrix}$$

Here the sub-matrices $J_1 \in \mathbb{C}^{c_1 \times c_1}, \dots, J_{l_{2\pi}} \in \mathbb{C}^{c_{l_{2\pi}} \times c_{l_{2\pi}}}$ correspond to the eigenvalues of unit modulus, ordered according to increasing frequency $\omega_k \in [0, 2\pi)$ and the matrix J_{st} accounts for the eigenvalues with absolute value smaller than 1. It follows directly from the results in Bauer and Wagner (2002b) that the process is MFI(1), if all matrices J_k are of the form $J_k = z_k I_{c_k}$. This will henceforth be assumed. Partition also $B = \left[B_1', \ldots, B_{l_{2\pi}}', B_{st}'\right]'$ and $C = [C_1, \ldots, C_{l_{2\pi}}, C_{st}]$.

We have already discussed that real valuedness of y_t imposes a number of restrictions on the system matrices and the eigenvalues of A. Specifically it holds that for each unit root z_k with $\omega_k \in (0, \pi)$ there exists an index k', such that $\omega_{k'} = \bar{\omega}_k$ and moreover $c_k = c_{k'}$. For the corresponding blocks of the system matrices $B_{k'} = \bar{B}_k$ and $C_{k'} = \bar{C}_k$ hold. Taking these restrictions into account, an observationally equivalent representation of the following format, where the blocks corresponding to pairs of complex conjugate are grouped together, exists:

$$A = \left[\begin{array}{cccc} J_{c_1}(z_1) & 0 & \ldots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & J_{c_l}(z_l) & 0 \\ 0 & \ldots & 0 & J_{st} \end{array} \right], C = \left[\begin{array}{cccc} C_1(z_1) & \ldots & C_l(z_l) & C_{st} \end{array} \right], B = \left[\begin{array}{c} B_1(z_1) \\ \vdots \\ B_l(z_l) \\ B_{st} \end{array} \right] \tag{8}$$

Again $z_1, \ldots, z_l$ denote the l distinct unit roots with frequency in $[0, \pi]$. The matrices $J_k(z_k), B_k(z_k)$ and $C_k(z_k)$ are of the following form. For $z_k \notin \{1, -1\}$ we obtain

$$J_k(z_k) = \left[\begin{array}{cc} z_k I_{c_k} & 0^{c_k \times c_k} \\ 0^{c_k \times c_k} & \bar{z}_k I_{c_k} \end{array} \right], B_k(z_k) = \left[\begin{array}{c} B_k \\ \bar{B}_k \end{array} \right], C_k(z_k) = \left[\begin{array}{cc} C_k & \bar{C}_k \end{array} \right]$$

Transforming this sub-system with

$$T_k = \left[\begin{array}{cc} I_{c_k} & I_{c_k} \\ i I_{c_k} & -i I_{c_k} \end{array} \right]$$

one obtains a real valued representation of the sub-system $(T_k J_k(z_k) T_k^{-1}, T_k B_k(z_k), C_k(z_k) T_k^{-1})$ corresponding to the pair of complex conjugate roots given by:

$$\left[\begin{array}{cc} \cos\omega_k I_{c_k} & \sin\omega_k I_{c_k} \\ -\sin\omega_k I_{c_k} & \cos\omega_k I_{c_k} \end{array} \right], \left[\begin{array}{c} 2B_k^r \\ -2B_k^i \end{array} \right], \left[\begin{array}{cc} C_k^r & C_k^i \end{array} \right] \tag{9}$$

where superscript r denotes the real part of a complex number and superscript i is used to denote the imaginary part, e.g. $C_k = C_k^r + iC_k^i$. Thus, these blocks have the double size compared to the block-size when each root on the unit circle is treated or counted separately. For the real valued unit roots, the blocks of the system matrices are of the (unchanged) form:

$$J_{c_k}(1) = I_{c_k}, B_k(1) = B_k, C_k(1) = C_k \text{ for k: } z_k = 1$$

$$J_{c_k}(-1) = -I_{c_k}, B_k(-1) = B_k, C_k(-1) = C_k \text{ for k: } z_k = -1$$

Effectively, the state components are reordered such that the blocks corresponding to pairs of conjugate complex eigenvalues are showing up as one of the (possibly larger) blocks $(J_{c_k}(z_k), B_k(z_k), C_k(z_k))$. Denote by $x_{t,k}$ the state components corresponding to the block for unit root z_k and by $x_{t,st}$ the stationary components of the state. For $z_k \notin \{-1,1\}$ the dimension of $x_{t,k}$ is given by $2c_k$ and it is c_k for $z_k \in \{-1,1\}$. From this one directly obtains the following representation for y_t, where we assume for fixing notation that $z_1 = 1$ and $z_l = -1$ are present. If these unit roots are not present the corresponding matrices can be set to $B_1 = C_1 = 0$ and $B_l = C_l = 0$ below:

$$\begin{array}{rcl} y_t & = & C_1(z_1)x_{t,1} + \ldots + C_l(z_l)x_{t,l} + C_{st}x_{t,st} + Ds_t + \varepsilon_t \\ & = & C_1B_1(x_{1,1} + \sum_{i=1}^{t-1}\varepsilon_{t-i}) + \sum_{k=2}^{l-1}\left[C_kB_k(z_k^{t-1}x_{1,k} + \sum_{i=1}^{t-1} z_k^{i-1}\varepsilon_{t-i})\right] + \\ & & \sum_{k=2}^{l-1}\left[\bar{C}_k\bar{B}_k(\overline{z_k^{t-1}x_{1,k}} + \sum_{i=1}^{t-1}\bar{z}_k^{i-1}\varepsilon_{t-r})\right] + C_lB_l(x_{1,l} + \sum_{i=1}^{t-1}(-1)^{i-1}\varepsilon_{t-i}) + \\ & & Ds_t + k_{st}(z)\varepsilon_t \end{array} \qquad (10)$$

where $k_{st}(z) = I + \sum_{j=1}^{\infty} C_{st}J_{st}^{j-1}B_{st}z^j$ denotes the stable part of the transfer function and the second equation above is derived from the first one by using the state transition equation $x_{t+1} = Ax_t + B\varepsilon_t$ and the initial conditions for the nonstationary components of the state, partitioned accordingly to the structure of the partitioning of the state x_t. The above representation (10) of y_t directly shows the contribution of the stochastic trend (corresponding to unit root $z = 1$) and the stochastic cycles corresponding to the (pairs of complex conjugate) unit roots $z_2, \ldots, z_l$. In this sense representation (10) constitutes a generalization of Granger's representation theorem to MFI(1) ARMA processes. The contribution to y_t stemming from the unit root z_k (and its complex conjugate $\bar{z}_k$) is given by:

$$C_kB_k\sum_{i=1}^{t-1} z_k^{i-1}\varepsilon_{t-i} + \bar{C}_k\bar{B}_k\sum_{i=1}^{t-1}\bar{z}_k^{i-1}\varepsilon_{t-i} + C_kB_kz_k^{t-1}x_{1,k} + \overline{C_kB_kz_k^{t-1}x_{1,k}}$$

It is the sum of the conjugate complex stochastic cycles at frequency ω_k and $\bar{\omega}_k$ respectively and the effects of the initial values

$$C_kB_kz_k^{t-1}x_{1,k} + \overline{C_kB_kz_k^{t-1}x_{1,k}} = 2\left[\mathcal{R}(C_kB_kx_{1,k})\cos(\omega_k(t-1)) - \mathcal{I}(C_kB_kx_{1,k})\sin(\omega_k(t-1))\right]$$

where $\mathcal{R}$ denotes the real part and $\mathcal{I}$ the imaginary part of a complex quantity. This expression clearly is real valued. Minimality implies (cf. Lemma 2 of Bauer and Wagner, 2002), that C_k and B_k are of full rank. For the MFI(1) case it is seen that the number of Jordan blocks (all of size one) corresponding to the eigenvalues z_k equals the number of common cycles corresponding to this unit root. This motivates the definition of the state space integration structure:

Definition 3 *The state space integration structure or state space unit root structure of a real valued MFI(1) process is given by* $\Omega = \{(\omega_1, c_1), \ldots, (\omega_l, c_l)\}$*, where* $z_k = e^{i\omega_k}$ *with* $\omega_k \in [0, \pi]$ *denotes the unit roots ordered according to increasing frequency and* c_k *denotes the number of common trends respectively common cycles corresponding to the unit root* z_k*.*

Remark 4 *If the restriction to real valued processes is omitted, the state space integration structure is to be defined incorporating all unit roots* $\omega_k \in [0, 2\pi)$ *separately. A similar remark also applies to the integration structure in Definition 1 and the definition of cointegration in Definition 2. In case that integration or cointegration is analyzed for only one (complex) root* z_k *with frequency in* $\omega_k \in [0, 2\pi)$*, we use the term complex integrated or complex cointegrated if the resulting time series are filtered with the complex filter* $(1 - z_k z)$ *only.*
If one wants to keep the analysis at a maximum level of generality, furthermore the cointegrating vectors β *and the dynamic cointegrating relationships* $\beta(z) = \beta_0 + \beta_1 z$ *can be allowed to have complex valued coefficients* β, β_0 *and* β_1*.*

Note that in order to ensure that also $k_{st}(z)$ and $x_{t,st}$ generate real valued output, similar restrictions as for the nonstationary part have to hold for the stable eigenvalues of A and the corresponding sub-blocks of B_{st} and C_{st}. Note, however, that due to the block-diagonal structure of the Jordan normal form J, the stable part of the transfer function is decoupled from the nonstationary part and thus (A_{st}, B_{st}, C_{st}) can be dealt with independently of the unit roots. Therefore, any real canonical form and parameterization for $k_{st}(z)$ can be used.
The system representation developed above allows to investigate the contribution of each unit root $\omega_k \in [0, 2\pi)$ to the output separately. Consequently it is also possible to consider (complex) cointegrating relationships that wipe out only the nonstationary contributions corresponding to one unit root of a pair of complex conjugate roots. The number of common cycles for a unit root z_k is given by the rank of $C_k B_k$, the same number of complex conjugate cycles is then also present for the unit root $\bar{z}_k$. As already mentioned above, for minimal representations it can be shown (cf. Lemma 2 in Bauer and Wagner, 2002b) that the rank of both B_k and C_k is equal to c_k, i.e. both matrices have full rank. Thus, there exists a matrix $C_k^\perp \in \mathbb{C}^{s \times r_k}$ with $0 \leq r_k = s - c_k \leq s$, such that $(C_k^\perp)' C_k^\perp = I_{r_k}$ and $(C_k^\perp)' C_k = 0$, i.e. $C_k^\perp$ spans the orthogonal complement to the space spanned by C_k. Now, multiplying y_t from the left by $(C_k^\perp)'$, using e.g. equation (10), one immediately sees that the columns of $C_k^\perp$ span the *complex* cointegrating space corresponding to the unit root z_k. It is obvious that

the cointegrating space to the complex conjugate root is given by $\overline{(C_k^\perp)} = (\bar{C}_k)^\perp$. Thus, the complex valued cointegrating space that wipes out the stochastic cycle corresponding to z_k and $\bar{z}_k$ is given by the intersection of the spaces spanned by $C_k^\perp$ and $(\bar{C}_k)^\perp$. For a vector β only contained in the span of $C_k^\perp$ but not in the complex conjugate space, the resulting series $\beta' y_t$ still contains the nonstationarities corresponding to the unit root at $\bar{z}_k$ and is complex valued.

For complex unit roots also *dynamic* cointegrating relationships may be present, these are linear polynomials in the backward shift operator $\beta(z) = \beta_0 + \beta_1 z$, with $\beta_0, \beta_1 \in \mathbb{R}^s$ or $\mathbb{C}^s$. The developed state space representation is also very revealing in showing why cointegrating relationships of this form may be present for MFI(1) processes. Look only at one term of representation (10) to obtain

$$(\beta_0' + \beta_1' z) C_k B_k \sum_{j=1}^{t-1} z_k^{j-1} \varepsilon_{t-j} = \beta_0' C_k B_k \varepsilon_{t-1} + \left[\beta_0' C_k z_k + \beta_1' C_k\right] B_k \sum_{j=1}^{t-2} z_k^{j-1} \varepsilon_{t-j-1}$$

Thus, dynamic complex cointegration at the unit root z_k occurs for

$$\left[\begin{array}{cc} \beta_0' & \beta_1' \end{array}\right] \left[\begin{array}{c} C_k z_k \\ C_k \end{array}\right] = 0 \tag{11}$$

Using the stacked notation $\beta(z) = \left[\begin{array}{cc} \beta_0' & \beta_1' \end{array}\right]'$, also the dynamic cointegrating relationships are found via orthogonality relationships over a space of dimension $2s$.

Also in the corresponding real valued system representation the dynamic cointegrating relationships can be recovered via orthogonality relationships: Denoting the real matrices as given in (9) by $(A_{k,\mathbb{R}}, B_{k,\mathbb{R}}, C_{k,\mathbb{R}})$ it follows that

$$(\beta_0 + \beta_1 z)' C_{k,\mathbb{R}} \sum_{i=1}^{t-1} A_{k,\mathbb{R}}^{i-1} B_{k,\mathbb{R}} \varepsilon_{t-i} = \beta_0' C_{k,\mathbb{R}} B_{k,\mathbb{R}} \varepsilon_{t-1} + \left(\beta_0' C_{k,\mathbb{R}} A_{k,\mathbb{R}} + \beta_1' C_{k,\mathbb{R}}\right) \sum_{i=1}^{t-2} A_{k,\mathbb{R}}^{i-1} B_{k,\mathbb{R}} \varepsilon_{t-i-1}$$

and thus the common cycle is eliminated, if and only if

$$\left[\begin{array}{cc} \beta_0' & \beta_1' \end{array}\right] \left[\begin{array}{c} C_{k,\mathbb{R}} A_{k,\mathbb{R}} \\ C_{k,\mathbb{R}} \end{array}\right] = 0 \tag{12}$$

which again is a simple orthogonality restriction, of the same type as (11) in the complex representation. In Bauer and Wagner (2002c) it is shown, that also for processes with higher integration orders, all higher order polynomial cointegration vectors of the form $\beta(z) = \sum_{j=0}^{q} \beta_j z^j$ can be found using similar orthogonality restrictions.

3 A Parameterization for MFI(1) Processes and an Example

In the previous section it has been shown that each solution to the system equations (1) with state space integration structure $\Omega = \{(\omega_1, c_1), \ldots, (\omega_l, c_l)\}$ is an MFI(1) process. In fact it can be shown that each rational MFI(1) process can be given a state space representation, see the following corollary, where we denote with $M_n(\Omega) \subset M_n$ the set of all rational transfer functions $k(z) \in M_n$, which correspond to state space integration structure Ω.

Corollary 1 *Let $y_t, t \in \mathbb{N}$ denote a real valued rational MFI(1) process. Then there exists a minimal order n and a state space realization $(A, B, C) \in S_n$, such that $y_t, t \in \mathbb{N}$ is generated by the state space equations (1) for some suitable initial value $x_1 \in \mathbb{C}^n$. Consequently there exists a state space integration structure Ω, such that $k(z) = \Pi(A, B, C) \in M_n(\Omega)$.*

`PROOF:` The corollary is a special case of results derived in Theorem 1 in Bauer and Wagner (2002b).

Thus, for each rational MFI(1) process state space representations exist. For estimation purposes identifiability has to be ensured, i.e. a unique representative of the class of observationally equivalent systems has to be selected. This is done by so called *canonical forms*. In the previous section the discussion showed that restricting the A-matrix to be (with possibly a specific ordering to obtain the blocks $J_k(z_k)$) in Jordan normal form leads directly to a very revealing representation of the system dynamics, e.g. with respect to the contributions attributable to the various unit roots, see (10). However, restricting the A-matrix to be in the discussed format does not achieve identification. Look for simplicity of the argument again only at one block corresponding to one unit root, (J_k, B_k, C_k). Then it holds for any nonsingular matrix T_k, that the (sub)system $(J_k = T_k J_k T_k^{-1}, T_k B_k, C_k T_k^{-1})$ is observationally equivalent and also has its A-matrix in Jordan normal form. Hence, further restrictions have to be imposed on either the B- or the C-matrix, or both. The canonical representation, developed in Bauer and Wagner (2002b), places the further required restrictions only on the blocks C_k, which are, as has been seen in the previous section, linked via orthogonality relationships to the cointegrating spaces. A canonical representation is obtained by requiring the matrices C_k to be orthonormal, i.e. $C_k' C_k = I_{c_k}$ and *positive lower triangular*. A complex matrix C is called positive lower triangular (p.l.t.), if it is of the form (see Ober, 1996 for a definition and properties):

$$C = \begin{bmatrix} 0 & 0 & \dots & 0 \\ c_{j_1,1} & \vdots & & \vdots \\ \vdots & 0 & & \vdots \\ \vdots & c_{j_2,2} & \ddots & 0 \\ \vdots & \vdots & & c_{j_m,m} \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix} \tag{13}$$

Here $c_{j_i,i} > 0, i = 1, \dots, m, 1 \le j_i < j_{i+1} \le n, i = 1, \dots, m-1$, is without restriction of generality real valued. I.e. the first non-zero element in each column of C is positive, which explains the name *positive* lower triangular.

For the canonical form just described, a parameterization follows immediately. The block-diagonal structure implies that the nonstationary and the (asymptotically) stationary part can be treated completely separately. For given state space integration structure Ω, the part of the A-matrix corresponding to the eigenvalues of modulus one is completely determined. Therefore there are no free parameters corresponding to the nonstationary part of A. With respect to the blocks $C_k, k = 1, \dots, l$, the restriction to orthonormality and to p.l.t. format have to be taken into account. The latter introduces additional integer parameters to describe the structure of the p.l.t. matrices, see (13). In Appendix B one parameterization for orthonormal p.l.t. matrices is described. This specific parameterization is based on stereographic projections. All entries in the matrices $B_k, k = 1, \dots, l$ are free parameters. Here minimality implies that the matrices B_k are of full rank. For the stationary part, $k_{st}(z)$, any suitable parameterization can be employed, e.g. echelon parameters.

Let θ denote a multi-index comprising Ω, the state space integration structure, the indices describing the p.l.t. structure of the matrices C_k ($\theta_{plt} = \theta_{plt}(\Omega)$) and indices necessary to describe $k_{st}(z)$, the stable part of the transfer function (θ_{st}). Note that θ_{plt} depends upon Ω. Denote by $\Theta(\Omega)$ the set of all feasible parameters $\theta = [\Omega, \theta_{plt}, \theta_{st}]$ corresponding to a given Ω. The set M_n cannot be parameterized continuously, see Hazewinkel and Kalman (1976). Hence, continuous parameterizations have to be based on a partitioning of M_n. In our case we partition $M_n = \bigcup_\theta M_n^\theta$, where M_n^θ denotes the set of all transfer functions $k(z) \in M_n$, which correspond to the index θ. When concentrating on the nonstationary part, it is convenient to partition M_n only according to Ω. Then $M_n = \bigcup_\Omega M_n(\Omega)$, where $M_n(\Omega) = \bigcup_{\theta \in \Theta(\Omega)} M_n^\theta$. The following corollary now presents some results concerning a canonical form and a parameterization of the set $M_n^\theta(\Omega)$.

Corollary 2 *For each transfer function $k(z) \in M_n(\Omega), \Omega = \{(\omega_1, c_1), \dots, (\omega_l, c_l)\}$, which generates a real rational multiple frequency I(1) output process, there exists a unique state space system (A, B, C), such that $\Pi(A, B, C) = k(z)$ and such that*

- *(A, B, C) is in the form (8), where (J_{st}, B_{st}, C_{st}) is e.g. in the echelon canonical form (see Hannan and Deistler, 1988, Chapter 2),*
- *where $C_k' C_k = I_{c_k}$ and C_k is p.l.t. for $k = 1, \dots, l$.*

Collect in $\theta = [\Omega, \theta_{plt}, \theta_{st}]$ for given Ω the integration structure, a set of integers specifying the structure of the p.l.t. matrices $C_k, k = 1, \dots, l$ and the integer parameters for the stable part of the transfer function $k_{st}(z)$. Let $c_k(z_k) = c_k$ for $z_k = \pm 1$ and $c_k(z_k) = 2c_k$ else. Then a set of parameters is given by

- *$\bar{\tau}_{st} \in \mathbb{R}^{d_{st}(\theta)}$, the parameters for the stationary sub-system $\Pi(A_{st}(\tau), B_{st}(\tau), C_{st}(\tau_{st})) = k_{st}(z)$ e.g. obtained from the echelon canonical form.*
- *$\tau_B \in \mathbb{R}^{(\sum_{k=1}^{l} c_k(z_k)) \times s}$, the parameters for the matrices $B_1 \in \mathbb{C}^{c_1(z_1) \times s}, \dots, B_l \in \mathbb{C}^{c_l(z_l) \times s}$.*
- *parameters τ_u for the matrices $C_1(\tau_u) \in \mathbb{C}^{s \times c_1}, \dots, C_l(\tau_u) \in \mathbb{C}^{s \times c_l}$, such that $C_k(\tau_u)' C_k(\tau_u) = I_{c_k}$ and $C_k(\tau_u)$ is p.l.t. for $k = 1, \dots, l$ and where the p.l.t. structure is determined by θ. One possibility to derive these parameters τ_u is presented in Lemma 10 given in Appendix B.*

Let ρ denote the mapping attaching the parameter $\tau \in \mathbb{R}^{d_\theta}$ to a transfer function $k(z) \in M_n^\theta(\Omega)$. Further let $T(\theta) = \rho(M_n^\theta(\Omega))$. Then, for each $\tau = [\tau_u', \tau_B', \bar{\tau}_{st}']' \in T(\theta)$ according to the above restrictions, let

$$
\begin{aligned}
A(\tau) &= diag(J_{c_1}(z_1), \dots, J_{c_l}(z_l), A_{st}(\bar{\tau}_{st})) \\
B(\tau) &= [B_1(z_1, \tau)', \dots, B_l(z_l, \tau)', B_{st}(\tau)']', \\
C(\tau) &= [C_1(z_1, \tau), \dots, C_l(z_l, \tau), C_{st}(\tau)]
\end{aligned}
$$

Then $\Pi(A(\tau), B(\tau), C(\tau)) \in \overline{M_n^\theta(\Omega)}$ and the subset $T^\circ(\theta) \subset \overline{T(\theta)}$, such that $(A(\tau), B(\tau), C(\tau))$ is minimal, is open and dense in $\overline{T(\theta)}$. The mapping $\varphi : T(\theta) \to \bar{M}_n^\theta(\Omega), \tau \mapsto k(z, \tau)$ is a parameterization of $M_n^\theta(\Omega)$, i.e. it is bijective and furthermore it is also continuous. Here $T(\theta)$ is endowed with the Euclidean metric and M_n with the so called pointwise topology (see Hannan and Deistler, 1988).

PROOF: The corollary is a special case of Theorem 2 in Bauer and Wagner (2002b).

For given state space integration structure Ω and given structure index θ, the above result provides a parameterization of the set of all transfer functions $M_n^\theta(\Omega)$. Note again that in the parameterization the nonstationary and the stationary part are decoupled. This implies that the stationary part can essentially be analyzed using well known techniques for stationary processes. Parameterizing the stationary part with echelon parameters, as formulated in the corollary, is only one possibility, any commonly used parameterization with the usual properties may be employed. Also the parameterization of the matrices C_k can be chosen by the user, given that some properties like differentiability are fulfilled (see the detailed discussion of Theorem 2 in Bauer and Wagner, 2002b). The parameterization presented in Appendix B is just one possibility in this respect.

The parameterization presented in the corollary takes into account the restriction that y_t is assumed to be real valued. Hence, parameters only have to be found for one of the blocks corresponding to pairs of conjugate complex unit roots (i.e. for B_k and C_k). If one wants to drop the restriction that y_t is real valued the results holds true with the obvious change that the blocks to each of the unit roots have to be parameterized. Note also that the parameterization is described in terms of real valued parameters.

The importance of the topological properties of the parameterization is discussed in detail in Bauer and Wagner (2002b). Let us therefore only briefly discuss the main issues: On the sets $T^\circ(\theta) \subset T(\theta)$ the mapping φ is a homeomorphism and twice continuously differentiable. This continuity property implies that convergence to the true transfer function in the set $M_n^\theta(\Omega)$ implies also convergence of the parameter vectors. Thus, it is sufficient to derive consistency at the level of transfer functions, which is the approach followed in the next section in Theorem 1. If furthermore the true parameter vector corresponds to an interior point of $T(\theta)$, linearization techniques can be employed in deriving asymptotic distributions of parameter estimates (see Theorem 2 below). Let us also note the fact that there exists a θ such that $M_n^\theta(\Omega)$ is 'generic' in $M_n(\Omega)$, i.e. contains almost all transfer functions. This choice of the index $(\theta_{plt}, \theta_{st})$ can be chosen as a starting point for optimization in the likely case that the integer parameters collected in θ, for given Ω are unknown. For the stationary part the generic choice for the integer parameters depends upon the chosen parameterization, and for the nonstationary part it is easily seen that the generic case corresponds to matrices C_k with entries along the main diagonal. Within the estimation procedure the suitability of the chosen index has to be tested and a re-specification may turn out to be advisable.

To illustrate the results discussed so far, this section is closed with a small example. Consider a model (e.g. for quarterly observations) with unit roots at $\pm 1, \pm i$ and output dimension $s = 4$. For simplicity any short-run dynamics are neglected (i.e. $k_{st}(z) = I_s$):

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & i & 0 \\ 0 & 0 & 0 & -i \end{bmatrix}, B = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -i \\ 0 & 0 & 1 & i \end{bmatrix},$$

$$C = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{3} & 1/2 & 1/2 \\ 1/\sqrt{2} & 1/\sqrt{3} & -1/2 & -1/2 \\ 0 & 0 & 1/2 - i/2 & 1/2 + i/2 \\ 0 & 1/\sqrt{3} & 0 & 0 \end{bmatrix}$$

This system is already given in the discussed (complex) canonical form, each of the blocks of C, in this case the four columns, is p.l.t. and has norm 1. The state space integration structure is given by $\Omega = \{(0, 1), (\frac{\pi}{2}, 1), (\pi, 1)\}$. Note that although the matrices A, B and C contain complex numbers, the output generated by this system is real valued. This can be verified by computing the impulse response coefficients $K_j = CA^{j-1}B$, which are real valued for all $j \geq 1$.

Parameters for the system are given next: The real valued parameters for B are $\tau_B = [B_1, B_2, B_3^r, B_3^i]'$, where B_k denotes the k-th row of B. The parameter vector τ_C consists of parameters for the columns of C in the following way, using the parameterization proposed in Appendix B. The columns C_1 and C_2 are real valued, hence the stereographic projection in $\mathbb{R}^4$ leads to $\tau_{C,1} = [(2 + \sqrt{2})^{-1}, 0, 0]'$ and $\tau_{C,2} = [(3 + \sqrt{3})^{-1}, 0, (3 + \sqrt{3})^{-1}]'$. Columns 3 and 4 corresponding to complex conjugate roots are linked via complex conjugacy. For these the real valued parameter vector is obtained from performing the stereographic projection in $\mathbb{R}^7$, stacking the real and the imaginary part of the third column of C, where the first component of the imaginary part is restricted to be zero and hence can be omitted. This results in $\tau_{C,3} = [-1/3, 1/3, 0, 0, -1/3, 0]'$. Summing up in this example the parameter space $T(\theta)$ is of dimension 28, 16 dimensions from B and 12 from C. Additionally $T(\theta)$ is restricted to fulfill that $\|\tau_{C,k}\| \leq 1, k = 1, 2, 3$ and the rows of B have to be different from 0 (due to minimality). The structure index $\theta_{plt} = [1, 1, 1]$, which indicates (the generic case) that the first entry in each column of C is not equal to 0, θ_{st} and $\bar{\tau}_{st}$ are empty in this example.

As discussed already in the previous section, the cointegrating spaces can directly be seen from the canonical system representation. The complex valued cointegrating spaces corresponding to the *four* different unit roots are given by the ortho-complements in $\mathbb{C}^4$ of the respective columns of $C = [C_1, C_2, C_3, C_4]$. For the unit roots 1 and -1 a real valued basis of the respective cointegrating spaces is directly found and we obtain: $C_{1,\mathbb{C}}^{\perp} = \{(1,-1,0,0)', (0,0,1,0)', (0,0,0,1)'\}$, $C_{2,\mathbb{C}}^{\perp} = \{(0,0,1,0)', (1,-1,0,0)', (1,0,0,-1)'\}$. For the unit roots at $\pm i$ we find the (complex conjugate) cointegrating spaces to be $C_{3,\mathbb{C}}^{\perp} = (\bar{C}_{4,\mathbb{C}})^{\perp} = \{(1,1,0,0)', (0,0,0,1)', (1+3i, 0, 1-2i, 0)'\}$. Here, as expected, complex quantities appear. The intersection of the cointegrating spaces corresponding to the roots $\pm i$ is spanned by $\{(0,0,0,1)', (1,1,0,0)'\}$, where now again directly a real valued basis is found. From the above spaces one sees, as the intersection of all four spaces is 0, that there exists no cointegrating relationship that wipes out all nonstationarities. There exist however linear combinations that wipe out two or three unit roots: $\dim(C_{1,\mathbb{C}}^{\perp} \cap C_{2,\mathbb{C}}^{\perp}) = 2$ and $\dim(C_{1,\mathbb{C}}^{\perp} \cap C_{3,\mathbb{C}}^{\perp} \cap C_{4,\mathbb{C}}^{\perp}) = 1$. I.e. there exist two linear independent cointegrating relationships that wipe out the unit roots at ± 1, and one cointegrating relationship that wipes out the unit roots at $1, \pm i$ (given by $(0,0,0,1)'$). Thus, in these linear combinations of the output only the complex pair of unit roots at $\pm i$ and the unit root -1 respectively are present.

Let us finally turn to discuss also a real valued system representation, which is obtained by transforming the sub-system corresponding to the pair of unit roots $\pm i$, compare (9):

$$A_{\mathbb{R}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{bmatrix}, B_{\mathbb{R}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix},$$

$$C_{\mathbb{R}} = \begin{bmatrix} 1/\sqrt{2} & 1\sqrt{3} & 1/2 & 0 \\ 1/\sqrt{2} & 1/\sqrt{3} & -1/2 & 0 \\ 0 & 0 & 1/2 & -1/2 \\ 0 & 1/\sqrt{3} & 0 & 0 \end{bmatrix}$$

As only the sub-system corresponding to the roots $\pm i$ is transformed, only the corresponding sub-blocks of (A, B, C) have changed. E.g. it is obvious to see that the third and fourth column of $C_{\mathbb{R}}$ are the real respectively the imaginary part of C_3 from the complex canonical representation. Statements concerning the real valued cointegrating spaces follow immediately. The cointegrating spaces corresponding to the roots 1 and -1 coincide with the ones derived above, as there already a real valued basis has been given. A difference to the above considerations is that now only real valued linear combinations of the basis vectors are admitted. The real valued cointegrating space to $\pm i$ is given by $C_{3,4,\mathbb{R}}^{\perp} = \{(1,1,0,0)', (0,0,0,1)'\}$.

It is the ortho-complement in $\mathbb{R}^4$ to columns three and four of $C_{\mathbb{R}}$ or equivalently the intersection of $C_{3,\mathbb{C}}^{\perp} \cap (C_{4,\mathbb{C}}^{\perp} = (\bar{C}_{3,\mathbb{C}})^{\perp})$ with $\mathbb{R}^4$.

The very last point in the discussion of the cointegration properties of the system is to find the dynamic cointegrating relationships, compare (12). The matrices corresponding to the roots $\pm i$ are given by:

$$A_{3,\mathbb{R}} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, B_{3,\mathbb{R}} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, C_{3,\mathbb{R}} = \begin{bmatrix} 1/2 & 0 \\ -1/2 & 0 \\ 1/2 & -1/2 \\ 0 & 0 \end{bmatrix},$$

Hence, the solution to equation (12) is a 6-dimensional space, using again the stacked notation $\beta(z) = [\beta_0', \beta_1'] \in \mathbb{R}^8$. Hence, six basis vectors have to be given. Four basis vectors are derived from the already presented static cointegrating vectors $C_{3,4,\mathbb{R}}^{\perp}$ by taking the vectors as well as lagging them, e.g. $\beta_0 = [0,0,0,1]', \beta_1 = 0$ and vice versa. The remaining two linearly independent vectors are given by $[0,0,1,0]' + z[0,2,1,0]'$ and $[1,3,1,0]' + z[0,0,-1,0]'$.

4 Pseudo Maximum Likelihood Estimation

After having laid the necessary foundations in the previous section, we can now turn to the central aim of the paper, namely to pseudo maximum likelihood estimation of the parameters introduced in the previous section for rational MFI(1) processes in the developed state space representation.

We first have to clarify the meaning of pseudo maximum likelihood estimation. To this end start with the assumption that ε_t is Gaussian distributed. Under this assumption it immediately follows that also $y_t, t \in \mathbb{N}$ is Gaussian distributed, since x_1 is assumed to be constant. Furtheron let (A^0, B^0, C^0) and (τ^0, D^0) (and Σ^0) respectively refer to the true system, whereas the quantities (A, B, C) and (τ, D) (and Σ) refer to any system in the considered class of systems. Define $\varepsilon_t(\tau, D) = \sum_{j=0}^{t-1} K_j^-(\tau)(y_{t-j} - Ds_{t-j})$, where $k^{-1}(z,\tau) = \sum_{j=0}^{\infty} K_j^-(\tau) z^j$ denotes the power series expansion of the inverse transfer function $k^{-1}(z) = I_s - zC(I_n - (A-BC)z)^{-1}B$, which converges uniformly for $|z| \leq 1$ due to the strict miniphase assumption. It now follows from the definition of the inverse of the transfer function, that $\sum_{j=0}^{i} K_j^-(\tau) K_{i-j}(\tau) = \delta_{0,i}$. Consequently it holds that

$$\varepsilon_t(\tau^0, D^0) = \varepsilon_t + \sum_{j=0}^{t-1} K_j^-(\tau^0)[C(\tau^0)A(\tau^0)^{t-j-1}x_1]$$

and thus as expected $\varepsilon_t(\tau, D)$ is an estimate of the innovation sequence. It follows that also $\varepsilon_t(\tau, D)$ is Gaussian distributed, as it is given by an inverse filtering of y_t, which is then Gaussian itself. Therefore $-2/T$ times the log likelihood function as a function of (τ, D, Σ) can be written as

$$\mathcal{L}_T(\tau, D, \Sigma; y_1, \ldots, y_T) = \log\det\Sigma + \frac{1}{T}\sum_{t=1}^{T} \varepsilon_t(\tau, D)'\Sigma^{-1}\varepsilon_t(\tau, D)$$

In case that the sequence ε_t is not Gaussian distributed, this function is not the (log) likelihood function. However, the Gaussian likelihood still constitutes a useful criterion to optimize, hence the name *pseudo maximum likelihood* estimation. Especially we will show below that maximizing the Gaussian likelihood leads under the standard assumptions (formulated in Section 2) on the noise sequence ε_t to consistent estimates.

With respect to the deterministic variables s_t, the results will be presented for two scenarios: In the first case the situation, where it is known, that there are no deterministic components s_t present in the process, is considered. This case is equivalent to assuming $D^0 = 0$ and moreover the restriction $D = 0$ is imposed. In the second scenario it is assumed, that the coefficient matrix D corresponding to correctly specified s_t is included in the estimation, i.e. all deterministic components present in the process are included in the estimation. Note, that it follows e.g. from (5) that the effect of the initial state is contained in the term $CA^{t-1}Bx_1 = 2\sum_{j=1}^{l}\mathcal{R}(C_j z_j^{t-1}B_j x_{1,j}) + C_{st}A_{st}^{t-1}B_{st}x_{1,st}$. Here the first summand fulfills that $(1 - z_k z)z_k^{t-1}x_{1,k} = (z_k^{t-1} - z_k^{t-1})x_{1,k} = 0$ and can thus be alternatively be attributed to the corresponding coordinate $s_{t,k}^1$ of s_t. Hence a nonidentifiability exists with regard to the effects of the initial state in the coordinates corresponding to the unit roots and the effects of the deterministic terms contained in $D_k s_{t,k}^1$, where D_k denotes the column of D corresponding to $s_{t,k}^1$. In order to make the representation unique, we assume $C_k' D_k = 0$. From the arguments given above it follows, that this is no restriction of generality. The second summand above converges to zero at an exponential rate and therefore does not influence the asymptotic behavior of the estimates. Hence, in the first scenario one effectively assumes, that there are no deterministic terms present in the process, which cannot be interpreted as the effects of nonzero initial conditions for the state.

The (pseudo) likelihood just defined, allows to solve for an explicit expression with respect to D. I.e. the likelihood can be concentrated and one obtains $\hat{D} = D(\tau, \Sigma)$: Consider the derivative of $\mathcal{L}_T$ with respect to the (i, j)-th entry in D, which is denoted as $\partial\mathcal{L}_T$ for simplicity of notation:[1]

$$\partial \mathcal{L}_T(\tau, D, \Sigma; y_1, \ldots, y_T) = -\frac{2}{T}\sum_{t=1}^{T}(k^{-1}(z,\tau)E_{i,j}s_t)'\Sigma^{-1}\left[k^{-1}(z,\tau)(y_t - Ds_t)\right] \tag{14}$$

Here $E_{i,j}$ denotes a matrix of zeros except at the (i,j)-th entry, which is equal to unity. Therefore $E_{i,j}s_t = e_i s_{t,j}$, where e_i denotes the i-th vector of the canonical basis and $s_{t,j}$ denotes the j-th component of s_t. Consider $k^{-1}(z,\tau)s_{t,j}$ more closely: $k^{-1}(z,\tau)s_{t,j} = k^{-1}(\overline{z_j},\tau)s_{t,j} + (1 - z_j z)\tilde{k}_j(z,\tau)s_{t,j}$. Since $(1 - z_j z)s_{t,j} = 0$ by construction of s_t, it follows that $k^{-1}(z,\tau)s_{t,j} = k^{-1}(\overline{z_j},\tau)s_{t,j}$. Using this, equation (14) can be modified to

$$k^{-1}(\overline{z_j},\tau)'\Sigma^{-1}\frac{1}{T}\sum_{t=1}^{T}(k^{-1}(z,\tau)y_t)\overline{s_{t,j}} = k^{-1}(\overline{z_j},\tau)'\Sigma^{-1}\sum_{r=1}^{m}\left(\frac{1}{T}\sum_{t=1}^{T}s_{t,r}\overline{s_{t,j}}\right)k^{-1}(\overline{z_r},\tau)D_r \tag{15}$$

with D_r denoting the r-th row of D. The above equation (15) holds for all $j = 1,\ldots,m$ and consequently defines $\hat{D}$ as the solution to the equations. Note that $k^{-1}(\bar{z}_j,\tau)$ is singular, hence the solution to the set of equations (15) can only be found in e.g. the least squares sense. It will be shown in the course of the proof of Theorem 1 that indeed a solution exists. Using this solution as an estimated $\hat{D} = D(\tau,\Sigma)$, the pseudo likelihood can be concentrated to $\mathcal{L}_T(\tau,\hat{D},\Sigma; y_1,\ldots,y_T)$. In case no deterministic terms are included in the estimation, $\hat{D} = 0$. The next step is to minimize the function $\mathcal{L}_T(\tau,\hat{D},\Sigma; y_1,\ldots,y_T)$ with respect to the innovation variance Σ (where the subsequent analysis has to provide a justification for the validity of this step) to obtain the minimum $\hat{\Sigma} = T^{-1}\sum_{t=1}^{T}\varepsilon_t(\tau,\hat{D})\varepsilon_t'(\tau,\hat{D})$. Hence, after these two concentration steps, the concentrated likelihood is up to an additive constant given by:

$$L_T(\tau; y_1,\ldots,y_T) = \mathcal{L}_T(\tau,\hat{D},\hat{\Sigma}; y_1,\ldots,y_T) = \log\det\{T^{-1}\sum_{t=1}^{T}\varepsilon_t(\tau,\hat{D})\varepsilon_t'(\tau,\hat{D})\} \tag{16}$$

Given the multi-index θ the pseudo maximum likelihood estimate is obtained by maximizing (16) over a compact parameter set $S(\theta) \subset T(\theta)$ as

$$\hat{\tau} = \arg\min_{\tau\in S(\theta)} L_T(A(\tau), B(\tau), C(\tau); y_1,\ldots,y_T)$$

The above minimization problem is solved over a subset of a real vector space, which can be tackled with any general purpose optimization algorithm, e.g. the Gauss-Newton algorithm or other gradient based methods. For the applicability of this type of optimization procedures

[1] Note that in general D is complex valued, hence differentiation takes place with respect to complex quantities. The required changes if D is parameterized using real valued parameters (to parameterize the real and imaginary part respectively) are obvious.

it is necessary that the parameterization possesses some basic properties such as differentiability, which have been derived in Bauer and Wagner (2002b) for the parameterization reviewed in the previous section.

The estimation procedure starts from a given multi-index θ which also comprises the given state space integration structure Ω. The inclusion of the integration structure is a drawback of the results.

The parameter vector τ is partitioned in $\tau = [\tau_u', \tau_{st}']'$ as follows: In τ_u the parameters corresponding to the matrix $C_u = [C_1(z_1, \tau), \ldots, C_l(z_l, \tau)]$ are collected. All other parameters are contained in $\tau_{st} = [\tau_B', \bar{\tau}_{st}']'$. The reason for this partitioning is the fact that the parameters τ_u, corresponding to the cointegrating spaces, are estimated super-consistently, and all elements of τ_{st} are estimated with the standard rate $T^{1/2}$. Let the parameter space according to τ_u be denoted as Θ_u. We have already stated before that the parameterization presented in Appendix B is only one choice for parameterizing orthonormal p.l.t. matrices. In order to keep the proof parameterization independent, Θ_u is chosen to be the set of all $s \times c$ matrices $C_u = [C_1(z_1), \ldots, C_l(z_l)]$, such that $C_k(z_k) \in \mathbb{C}^{s \times c_k(z_k)}$, where $C_k(z_k) = [C_k, \overline{C_k}]$ for $z_k \notin \{0, \pi\}$ and $C_k(z_k) = C_k$ else, such that C_k is p.l.t. and additionally $C_k'C_k = I_{c_k}$ holds. Consequently we will synonymously use C_u for τ_u. The topology in this set is induced by the following metric: The distance between $C_u, D_u \in \Theta_u$ is defined as $dist(C_u, D_u) = \max_{k=1,\ldots,l}\{\max(\|(I_s - C_{u,k}C_{u,k}')D_{u,k}\|, \|(I_s - D_{u,k}D_{u,k}')C_{u,k}\|)\}$. In this metric a sequence of matrices $C_{u,T}$ converges to C_u if and only if all subspaces spanned by the columns of $C_{u,k,T}$ converge to the subspace spanned by $C_{u,k}$.

Note that the results presented in this section also hold true if the parameterization presented in Appendix B is used. In this case Θ_u is a subset of $\mathbb{R}^{d_u}$ for some integer d_u (specified in Appendix B). The results derived below can easily be applied to different parameterizations of C_u, provided that the mapping attaching the parameters τ_u (in the chosen parameterization) to the matrices C_u is continuous and differentiable. A final advantage of chosing the set of matrices directly is the fact that the integer parameters θ_{plt} become obsolete. They only need to be introduced in the actual parameterization.

In the course of the proof restrictions are placed on the parameter vector. These restrictions are placed on the whole parameter vector and lead to possible dependence of the restrictions on the set of valid parameter vectors τ_{st} on τ_u. For a given $\tau_u \in \Theta_u$ denote with $\Theta_{st}^{\tau_u}$ the set of parameter vectors τ_{st} that fulfill the given restrictions. It follows from the formulation of the restrictions, that there exists a neighborhood Θ_u^0 around the true value τ_u^0 and a neighborhood Θ_{st}^0 of the true value τ_{st}^0, such that $\Theta_u^0 \times \Theta_{st}^0 \subset T(\theta)$.

Remark 5 *Also for the parameter space Θ_{st} a similar approach as for Θ_u could be chosen. $\tau_{st} = [\tau_B', \bar{\tau}_{st}']'$, where $\bar{\tau}_{st}$ is the parameter vector corresponding to $k_{st}(z)$. Now, it is possible to either assume that $\bar{\tau}_{st} \in \mathbb{R}^{d_{\overline{st}}}$ for some integer $d_{\overline{st}}$ corresponding to a particular neighborhood of the echelon canonical form (given that in θ the relevant indices correspond to the echelon canonical form). This choice makes $\Theta_{st}^{\tau_u}$ a subset of a real vector space, which can be endowed with e.g. the Euclidean metric. Alternatively it is possible to use $\bar{\tau}_{st} = k_{st}(z)$ in the set of all rational, stable transfer functions of McMillan degree exactly equal to $n - c$. This set is then endowed with the pointwise topology and the consistency proof could be given directly at the level of the transfer function. This choice together with the above choice for C_u would render the introduction of the structure index θ unnecessary, only the state space integration structure would remain to be given.*

Let $k^0(z) = k(z, \tau_u^0, \tau_{st}^0) = zC_u^0(I - zJ_u^0)^{-1}B_u^0 + k_{st}(z, \tau_{st}^0) \in M_n^\theta(\Omega)$ denote the true transfer function, where $\tau^0 = \left[(\tau_u^0)', (\tau_{st}^0)'\right]'$ denotes the true parameter vector and J_u^0 denotes, as already in Section 2, the matrix corresponding to the unit roots $z_k, k = 1, \ldots, l$ as specified in Ω. The notation $k(z, \tau) \in M_n^\theta(\Omega)$ is used to make the dependence on the backward-shift and the parameter vector explicit. Further denote by (A^0, B^0, C^0) the state space canonical form representation corresponding to $k^0(z)$ as provided in Corollary 2 (also for blocks of these matrices the superscript ^{0} will be used to denote the true quantities) and by $\hat{\tau}$ again the parameter vector optimizing $L_T(\tau)$ over a suitably restricted compact parameter set $S(\theta) \subset T(\theta)$. Then, the following result can be shown, whose proof in connection with some useful lemmata is given in Appendix A:

Theorem 1 (Consistency) *Let y_t be a rational MFI(1) process generated by a system of the form (1), where the white noise ε_t is a strictly stationary martingale difference sequence fulfilling the standard assumptions. The initial state is assumed to be constant.*
Let the true multi-index θ be known and let the unit root frequencies be denoted by $\omega_k \in [0, \pi]$ for $k = 1, \ldots, l$. Assume that $S(\theta) \subset T(\theta)$ is compact and that

$$\max_{k=1,\ldots,l} \{ \max_{\tau \in S(\theta)} \|k_{2,k}^{-1}(e^{i\omega_k}, \tau_u, \tau_{st})\| \} \quad < \quad \infty \tag{17}$$

where $k_{2,k}(z, \tau) = k(z, \tau) - zC_k(\tau)(I - ze^{i\omega_k}I)^{-1}B_k(\tau)$. Assume that $k(z, \tau)$ is uniformly stably invertible, i.e. $\max_{\tau \in S(\theta)} |\lambda_{max}(A(\tau) - B(\tau)C(\tau))| < 1 - \rho$, for some $\rho > 0$.

With respect to the deterministic variables s_t we investigate two cases: In the first case no deterministic terms are included in neither the true process nor the estimation (i.e. $D = D^0 = 0$). In the second case the true deterministic terms s_t are included in the estimation. In both cases the (pseudo) maximum likelihood estimate $[\hat{\tau}_u', \hat{\tau}_{st}']' \in S(\theta)$ obtained by maximizing $L_T(\tau)$ over $S(\theta)$ converges in probability to the true parameter vectors τ_u^0 and τ_{st}^0. Furthermore $T^\gamma dist(\tau_u, \tau_u^0) \to 0$ in probability for all $0 < \gamma < 1$, i.e. τ_u is estimated super-consistently.

For the second case the following asymptotic behavior occurs: If $(1 - z_k z)s_{t,i} = 0$ holds, for the i-th column of $\hat{D}$, then $C_k^0 (C_k^0)' \hat{D}_i \to 0, (I - C_k^0 (C_k^0)')\hat{D}_i \to D_i^0$, i.e. the components in the range space of C_k^0 converge to zero, whereas the components in the ortho-complement of the range space of C_k^0 are estimated consistently. If $s_{t,i}$ does not correspond to a unit root, then $\hat{D}_i \to D_i^0$.

Some of the assumptions formulated in Theorem 1 are stronger than the assumptions usually made in a stationary setting. Most notably the assumption concerning uniformly bounded zeros of the system and the bound on the maxima of the inverses stated in equation (17). The first assumption guarantees that for all systems with $\tau \in S(\theta)$ the value of the transfer function is bounded away from zero uniformly at the unit roots. The second bound is of a purely technical nature, and is possibly not a necessary condition. If one wants to perform misspecification analysis in order to e.g. develop test procedures for the unit roots, the above assumptions possibly need to be weakened: It has to be expected that zeros or poles of the likelihood estimate approach the unit circle when Ω is misspecified. This is, however, left for future research.

Concerning the rates of convergence note, as already mentioned above, that the parameter vector τ_u (consequently also C_u) is estimated super-consistently. This implies that also the cointegrating spaces are estimated super-consistently for all unit roots. The parameters for B_u and the parameters corresponding to $k_{st}(z)$ are estimated with the standard rate $T^{1/2}$.

The distributional result is referring to finite dimensional parameters. Hence, a specific parameterization for C_u has to be chosen. In the following we use the results of Corollary 1 and the parameterization for orthonormal p.l.t. matrices given in Appendix B. Also D has to be parameterized if it is estimated; otherwise it is restricted to zero. D contains, apart from the restriction that the output y_t is assumed to be real valued, only free parameters, and real valued parameters are given by parameterizing the real and imaginary part.

The derivation of the asymptotic distribution follows the usual scheme. It proceeds in two steps and is based on linearization arguments around the true parameter value τ^0. The first step is the derivation of the asymptotic distribution of the score vector and the second step is to derive convergence of the suitably normalized Hessian of the log likelihood function. The approach is inspired by Saikkonen (1995) and the proof of the theorem is given in Appendix A:

Theorem 2 (Asymptotic Distribution) *Let y_t be a real valued MFI(1) process generated by a system of the form (1). Let $S(\theta)$ denote the compact set of parameter vectors over which the pseudo likelihood is optimized. Let the assumptions of Theorem 1 hold. Let the true transfer function $k^0(z)$ correspond to τ^0, an interior point of $S(\theta)$. Denote by $\hat{\tau} = [\hat{\tau}_u', \hat{\tau}_{st}']'$ the optimizing vector of the pseudo likelihood L_T over $S(\theta)$, then if no deterministic terms s_t are neither included in the true process nor in the estimation,*

$$\begin{aligned} T(\hat{\tau}_u - \tau_u^0) &\stackrel{d}{\to} Z_u \\ T^{1/2}(\hat{\tau}_{st} - \tau_{st}^0) &\stackrel{d}{\to} N(0, V) \end{aligned}$$

where Z_u is a mixture of Brownian motions and $N(0,V)$ is a multivariate normal with mean zero and variance V. Detailed expressions for Z_u can be obtained from the expressions given in Lemma 8 and Lemma 9 in the proof contained in Appendix A.
If the deterministic terms Ds_t are included in the estimation, let $d = [vec(\mathcal{R}(D))', vec(\mathcal{I}(D))']'$ and let d^0 denote the corresponding true parameter value. Then

$$\sqrt{T}(\hat{d} - d^0) \stackrel{d}{\to} \mathcal{N}(0, V_D)$$

where $V_D \geq 0$. In this case in the limiting distribution of $T(\hat{\tau}_u - \tau_u^0)$ the Brownian motions $W_k(u)$, contained in the expression for Z_u, have to be replaced by the corresponding demeaned Brownian motions $W_k(u) - \int_0^1 W_k(u)du$.

Note that the results on the asymptotic distribution directly lead to hypotheses tests on the parameters in the usual way, this includes also tests on the cointegrating relations.

5 Summary and Conclusions

In this paper we have shown that the state space framework is very suitable for the analysis of MFI(1) processes. MFI(1) processes, short for multiple frequency I(1) processes, are processes that are integrated of order 1 at a finite set of unit roots spread over the unit circle. We have discussed and derived pseudo maximum likelihood estimates and their asymptotic distribution, where we allow for the presence of deterministic cycles and a constant. The term pseudo maximum likelihood refers to the maximization of the Gaussian likelihood for errors that are not necessarily normally distributed, but only have to fulfill the standard assumptions formulated in Section 2. The main advantages of the state space framework are the simplicity of the representation of the system dynamics and cointegration properties. The second advantage is the equivalence of state space systems to ARMA systems, by which we overcome the common limitation in the literature to study autoregressive processes only.

The estimation results are based on a state space canonical form and parameterization that are discussed in the context of MFI(1) processes in Section 3. The results presented in this section are a special case of more general representation results derived for systems with integer integration orders corresponding to the various unit roots in Bauer and Wagner (2002b). The presented canonical representation has a couple of convenient features. E.g. it allows to obtain the cointegrating spaces solely from the matrix C. All cointegrating relationships, both static and dynamic, can be found via simple orthogonality relationships. In the presentation we use complex quantities, this simplifies both the algebra and leads also to more directly interpretable results concerning the nonstationary contributions to the output attributable to the different unit roots. As interest in econometrics is commonly restricted to real valued output processes, the implications of this restriction on the system matrices are discussed in detail. Based on this discussion, for a simple example system both real and complex valued canonical representations are presented for illustrative purposes. The fact that the cointegrating relationships are found via orthogonality constraints holds completely analogously for both the real as well as the complex valued representation. In the real valued representation the interplay of the complex conjugate contributions stemming from conjugate complex unit roots is highlighted.

The parameters describing the cointegrating spaces, τ_u, are shown to be estimated superconsistently, whereas all other parameters are estimated at rate $T^{1/2}$. This difference, as expected, has implications for the respective limiting distributions. τ_u has a mixture of Brownian motions limiting distribution. The precise form of the limiting distribution depends upon the state space integration structure and upon the inclusion of deterministic terms in the estimation. The latter is a standard feature in the unit roots and cointegration literature. τ_{st} and, if included in the estimation, also D are asymptotically normally distributed.

The main drawback of the results in this paper is the assumption of a correctly specified state space integration structure in the likelihood optimization problem. This complicates the development of test procedures for the dimensions of the cointegrating spaces at the various unit root frequencies. As mentioned already in the introduction, in Bauer and Wagner (2002a) a partial remedy to this problem is presented. In that paper, formulated in the state space framework in the same way as presented in this paper, various tests for the dimension of the cointegrating space are presented for the I(1) case. Estimation and testing in that paper are based on so called subspace algorithms. These are computationally extremely cheap procedures that not only lead to test procedures but also lead to consistent initial estimates for an I(1) likelihood estimation as a special case of the results presented in this paper. An alternative might be to use VAR approximations with orders increasing with sample size in testing for the various unit roots, analogously to previous results of Saikkonen (1992) and Saikkonen and Luukkonen (1997) in the I(1) case. However, this is left as a topic for further research.

Also a couple of other important questions are still open: The canonical form is constructed for processes with higher integration orders. Hence, the derivation of estimates for the general case is an important task. It seems in principle possible to extend the pseudo maximum likelihood approach pursued here for the MFI(1) case to higher order integrated processes. A second open point is the inclusion of (linear) trends in the present framework.

Acknowledgements

We want to thank Manfred Deistler, Søren Johansen and Benedikt Pötscher for directing our attention to the problem and for useful discussions.

A Proofs

Proof of Theorem 1 (Consistency)

The proof follows a similar line of thought as illustrated in the example given in Saikkonen (1995, Section 5). The key property in Saikkonen's work is the continuous convergence of certain quantities, which has also been developed in Saikkonen (1993). Instead of Saikkonen's Condition 3.1. (1993, p.160) we will use the following uniform equicontinuity condition, that is later shown to hold for the required quantities:

Condition 1 (USE - Uniform Stochastic Equicontinuity) *A sequence $X_n(\tau), \tau \in \Theta$ is said to fulfill Condition USE, if for every sequence $\tau_n \to \tau$ and every $\epsilon > 0, \delta > 0$ and $\eta > 0$ there exists an integer $n(\epsilon, \eta, \delta)$ such that $\mathbb{P}\{\sup_{t \in B(\tau_n, \delta)} \|X_n(t) - X_n(\tau_n)\| > \epsilon\} \leq \eta\delta$ for $n \geq n(\epsilon, \eta, \delta)$. Here $B(x, r)$ denotes the open ball with center x and radius r.*

This condition ensures that the convergence is uniformly in the parameter space. In our special case of a compact parameter space we obtain the following consequence:

Lemma 1 *Assume that $X_n(\tau), \tau \in \Theta$ fulfills Condition USE, where Θ is compact. Further assume that for each fixed $\tau \in \Theta$ the sequence $X_n(\tau) \to 0$ in probability for $n \to \infty$. Then $\sup_{\tau\in\Theta} X_n(\tau) \to 0$ in probability for $n \to \infty$.*

PROOF: Fix $\epsilon > 0, \delta > 0$. Let $\tau_1, \dots, \tau_k$ denote k points, such that $\cup_{i=1}^k B(\tau_i, \delta)$ covers Θ. Due to the assumed compactness of Θ a finite cover exists. Since $X_n(\tau)$ fulfills Condition USE, there exists for each i an integer $n_i(\epsilon, \eta, \delta)$ such that the probability that the error $\sup_{\tau'\in B(\tau_i,\delta)} \|X_n(\tau') - X_n(\tau_i)\| > \epsilon/2$ is smaller than $\eta\delta$ for $n \geq n_i(\epsilon, \eta, \delta)$. If $\eta = \epsilon/(2k\delta)$ is chosen,

$$\begin{aligned}
\mathbb{P}\{\sup_{\tau\in\Theta} \|X_n(\tau)\| > \epsilon\} &\leq \mathbb{P}\{\sup_{\tau\in\cup B(\tau_i,\delta)} \|X_n(\tau)\| > \epsilon\} \\
&\leq \sum_{i=1}^k \mathbb{P}\{\sup_{\tau\in B(\tau_i,\delta)} \|X_n(\tau) - X_n(\tau_i) + X_n(\tau_i)\| > \epsilon\} \\
&\leq \sum_{i=1}^k \mathbb{P}\{\sup_{\tau\in B(\tau_i,\delta)} \|X_n(\tau) - X_n(\tau_i)\| > \varepsilon/2\} + \mathbb{P}\{\|X_n(\tau_i)\| > \epsilon/2\} \\
&\leq k\eta\delta + \sum_{i=1}^k \mathbb{P}\{\|X_n(\tau_i)\| > \epsilon/2\}
\end{aligned}$$

and this can be made arbitrarily small by choosing n large, since $k\eta\delta = \epsilon/2$ and $X_n(\tau_i) \to 0$ in probability. □

In order to establish the fulfillment of this property, a key input are some convergence results that are summarized in the following lemma. Note that in the following lemmata we refer to the complex quantities and accordingly the roots are allowed to vary in the interval $[0, 2\pi)$. By $x_t(z_k)$ we denote the state components corresponding to exactly one unit root z_k. This is not to be confused with $x_{t,k}$ which collects all components corresponding to pairs of complex conjugate unit roots.

Lemma 2 *Let $\omega_1, \dots, \omega_{l_{2\pi}}$ denote $l_{2\pi}$ distinct frequencies in $[0, 2\pi)$ and let ε_t be a martingale difference sequence fulfilling the standard assumptions with nonsingular innovation variance Σ^0. Further let $x_{t+1}(z_k) = z_k x_t(z_k) + B_k\varepsilon_t, x_1(z_k) = x_1, z_k = e^{i\omega_k}$ and $s_{t+1}(z_k) = z_k s_t(z_k), s_1(z_k) = 1$ for $k = 1, \dots, l_{2\pi}$. Here x_1 is a complex valued constant vector. Let $\delta_k = 1$ for $z_k = \pm 1$ and $\delta_k = \frac{1}{\sqrt{2}}$ else. Further let $\delta_k W_k(w), k = 1, \cdots, l_{2\pi}$ denote the weak limit of $T^{-1/2}\sum_{t=1}^{[Tw]} \overline{z_k}^t \varepsilon_t$, where $[Tw]$ denotes the integer part of Tw. $W_k(w)$ and $W_j(w)$ are independent for $j \neq k$. $W_k(\omega) = W_k^r(\omega) + iW_k^i(\omega)$, where $W_k^r(\omega)$ denotes the real part and $W_k(\omega)$ the complex part of the random variable, which are independent real valued random walks with variance Σ^0.*
Then the following statements hold:

i) $T^{-1}\sum_{t=1}^{T-j} \varepsilon_t\varepsilon_{t+j}' \to \delta_{0,j}\Sigma^0$ *in probability, where* $\delta_{0,j} = 1$ *for* $j = 0$ *and zero else.*

ii) $T^{-1}\sum_{t=1}^T x_t(z_k)\varepsilon_t' \xrightarrow{d} \delta_k^2 B_k \int_0^1 W_k(w)dW_k(w)' =: X(z_k)$
$T^{-1}\sum_{t=1}^T s_t(z_k)\varepsilon_t' \to 0$

iii) $T^{-2}\sum_{t=1}^T x_t(z_k)x_t(z_k)' \xrightarrow{d} \delta_k^2 B_k \int_0^1 W_k(w)W_k(w)'dwB_k' =: Z(z_k)$
$T^{-1}\sum_{t=1}^T s_t(z_k)s_t(z_k)' = 1.$

iv) *For* $z_k \neq z_j$ *it holds that* $T^{-2}\sum_{t=1}^T x_t(z_k)x_t(z_j)' \to 0$ *and*
$T^{-1}\sum_{t=1}^T s_t(z_k)s_t(z_j)' \to 0$ *in probability.*

v) $T^{-3/2}\sum_{t=1}^T x_t(z_k)s_t(z_k)' \xrightarrow{d} \delta_k \int_0^1 B_k W_k(w)dw =: Y(z_k)$
and $T^{-3/2}\sum_{t=1}^T x_t(z_k)s_t(z_j)' \to 0$ *in probability for* $z_k \neq z_j$.

PROOF: The proof of the lemma is in many parts a direct consequence of results obtained in Johansen and Schaumburg (1999), Lemma 5, Theorem 6 and Corollary 7. One difference is the inclusion of starting values, which however does not influence the asymptotic behavior. This follows, since the initial effects can be alternatively represented in the appropriate term $Ds_t(z_k)$.
Also the results concerning $s_t(z_j)$ are standard, except for the cross terms with $x_t(z_k)$. For $z_j \neq z_k$ the proof is analogous to the proof of equation (22) in Johansen and Schaumburg (1999) and for the case $z_j = z_k$ the continuous mapping theorem leads to the stated results. □

A difference to Johansen and Schaumburg (1999) is that we investigate the unit roots separately for the interval $[0, 2\pi)$. This is however only a notational extension, as of course $X(z_k) = \overline{X(\bar{z}_k)}$ holds for the complex unit roots and analogous results also apply to $Y(z_k)$ and $Z(z_k)$.
In the expressions for the pseudo likelihood function, terms that can be represented as filtered versions of the observations y_t show up, where the filters depend upon the parameter values. This necessitates to understand the convergence properties of estimated sample covariances of expressions of the form $l(z,\tau)x_t(z_k) = \sum_{j=0}^{t-1} L_j(\tau)x_{t-j}(z_k)$, where $l(z,\tau) = \sum_{j=0}^{\infty} L_j(\tau)z^j$ denotes a family of stable transfer functions parameterized by the parameter vector τ. The notation here indicates, that the summation is only performed for $t > 0$ or equivalently $x_t(z_k) = 0, t < 0$ is assumed. A family of transfer functions $l(z,\tau), \tau \in \Theta$ is called *uniformly stable*, if there exist constants $C < \infty, 0 < \rho < 1$, such that $\sup_{\tau\in\Theta} \|L_j(\tau)\| \leq C\rho^j$, i.e. the decay in the transfer function coefficients is exponential and uniform in the parameter set. For quantities of this form in the following lemma the asymptotic behavior is clarified and for each of the considered expressions *Condition USE* is established. The lemma parallels Theorem 4.2 in Saikkonen (1993, p.167) in which he establishes his Condition 3.1.

Lemma 3 *Let $l(z,\tau) = \sum_{i=0}^{\infty} L_i(\tau)z^i, k(z,\tau) = \sum_{i=0}^{\infty} K_i(\tau)z^i, \tau \in \Theta$ be two uniformly stable families of rational transfer functions of finite McMillan degrees less or equal to n, where it is always assumed that the transfer functions are of the correct dimensions. Let ε_t be a martingale difference sequence fulfilling the standard assumptions with nonsingular innovation variance Σ. Furthermore let $\omega_1, \dots, \omega_{l_{2\pi}}$ denote $l_{2\pi}$ distinct frequencies in $[0, 2\pi)$ and let $z_k = e^{i\omega_k}$. Furthermore $x_t(z_k)$, $s_t(z_k)$, $X(z_k)$, $Y(z_k)$ and $Z(z_k)$ are as defined in Lemma 2.*
The following asymptotic results hold for each fixed $\tau \in \Theta$

i) $T^{-1}\sum_{t=1}^{T} l(z,\tau)\varepsilon_t(k(z,\tau)\varepsilon_{t-i})' \to \sum_{r=0}^{\infty} L_{r+i}(\tau)\Sigma K_r(\tau)'$ *in probability for $i \geq 0$.*
$T^{-1}\sum_{t=1}^{T} l(z,\tau)s_t(z_k)(k(z,\tau)\varepsilon_t)' \to 0.$

ii) $T^{-1}\sum_{t=1}^{T} l(z,\tau)x_t(z_k)(k(z,\tau)\varepsilon_t)' \xrightarrow{d}$

$$l(\overline{z_k},\tau)X(z_k)k(\overline{z_k},\tau)' - l(\overline{z_k})\overline{z_k}B_k\Sigma\tilde{k}(0)' + \lim_{t\to\infty} \mathbb{E}\tilde{l}(z,\tau)B_k\varepsilon_{t-1}(k(z,\tau)\varepsilon_t)'$$

where $l(z,\tau) = l(\overline{z_k},\tau) + (1 - z_k z)\tilde{l}(z,\tau), k(z,\tau) = k(\overline{z_k},\tau) + (1 - z_k z)\tilde{k}(z,\tau).$

iii) $T^{-2}\sum_{t=1}^{T} l(z,\tau)x_t(z_k)(k(z,\tau)x_t(z_k))' \xrightarrow{d} l(\overline{z_k},\tau)Z(z_k)k(\overline{z_k},\tau)'.$

iv) $T^{-2}\sum_{t=1}^{T} l(z,\tau)x_t(z_k)(k(z,\tau)x_t(z_j))' \xrightarrow{d} 0$ *for $z_k \neq z_j$.*

v) $T^{-1}\sum_{t=1}^{T} l(z,\tau)s_t(z_k)(k(z,\tau)s_t(z_j))' \to l(\overline{z_k},\tau)k(\overline{z_k},\tau)'$ *for $z_k = z_j$ and to zero else.*

vi) $T^{-3/2}\sum_{t=1}^{T} l(z,\tau)x_t(z_k)(k(z,\tau)s_t(z_j))' \to l(\overline{z_k},\tau)Y(z_k)k(\overline{z_k},\tau)'$ *if $z_k = z_j$ and to zero else.*

All sequences in items i) to vi) fulfill condition USE.

PROOF: The proof of the lemma rests upon the results established in Lemma 2. Item i) is standard and its proof is therefore omitted. Analogously to the well known decomposition for the case $z_k = 1$, decompose $l(z,\tau) = l(\overline{z_k},\tau) + (1 - z_k z)\tilde{l}(z,\tau)$ for each τ and $|z_k| = 1$. The assumed uniform stability of $l(z,\tau)$ implies that also $\tilde{l}(z,\tau) = \sum_{j=0}^{\infty} \tilde{L}_j(\tau) z^j$ is a uniformly stable family of transfer functions. Using the decomposition we obtain:

$$\begin{array}{rcl} l(z,\tau)x_t(z_k) & = & \sum_{i=0}^{t-1} L_i(\tau)x_{t-i}(z_k) \\ & = & l(\overline{z_k},\tau)x_t(z_k) + \tilde{l}(z,\tau)(x_t(z_k) - z_k x_{t-1}(z_k)) \\ & = & l(\overline{z_k},\tau)x_t(z_k) + \tilde{l}(z,\tau)B_k\varepsilon_{t-1} + \tilde{L}_{t-1}(\tau)x_1(z_k) \end{array}$$

for $t \in \mathbb{N}$. Now, due to the fact that also $\tilde{L}_{t-1}(\tau)$ converges uniformly in $\tau \in \Theta$ at an exponential rate to zero, the last term in the above expression can be neglected. Then item ii) follows from

$$T^{-1}\sum_{t=1}^{T} x_t(z_k)(k(z,\tau)\varepsilon_t)' = T^{-1}\sum_{t=1}^{T} x_t(z_k)\varepsilon_t' k(\overline{z_k},\tau)' + T^{-1}\sum_{t=1}^{T} x_t(z_k)(\tilde{k}(z,\tau)(1-z_k z)\varepsilon_t)'$$

The first term in this sum converges to $X(z_k)k(\overline{z_k},\tau)'$ according to Lemma 2. The second term in this sum is equal to

$$T^{-1}\sum_{t=1}^{T} x_t(z_k)(\tilde{k}(z,\tau)\varepsilon_t)' - \overline{z_k}x_t(z_k)(\tilde{k}(z,\tau)\varepsilon_{t-1})' = \overline{z_k}\left(T^{-1}\sum_{t=1}^{T-1}\left[z_k x_t(z_k) - x_{t+1}(z_k)\right]\left(\tilde{k}(z,\tau)\varepsilon_t\right)'\right) + o_P(1)$$

where the $o_P(1)$ term is due to $T^{-1}x_T(z_k)(\tilde{k}(z,\tau)\varepsilon_T)'$. This term converges to $-\overline{z_k}B_k\mathbb{E}\varepsilon_t(\tilde{k}(z,\tau)\varepsilon_t)' \to -\overline{z_k}B_k\Sigma\tilde{k}(0,\tau)'$ for $t \to \infty$. Combining this with pre-multiplication of $x_t(z_k)$ with $l(\overline{z_k},\tau)$ then delivers the result. Items iii) and iv) can be shown using similar arguments. The proof of v) and vi) follows from $l(z,\tau)s_t(z_k) = l(\overline{z_k},\tau)s_t(z_k) + l(\overline{z_k},\tau)(1-z_k z)s_t(z_k)$ and $(1-z_k z)s_t(z_k) = 0$ for $t > 1$ and 1 for $t = 0$.

The fulfillment of Condition USE for the sequences considered in i) to vi) is left to be shown. For i) the claim follows from standard arguments for stationary processes. The difference for two parameter vectors (remembering that Condition USE is concerned with the behavior for $\tau_n \to \tau$) can be decomposed in two parts: One part depends only upon the parameter vectors but not on ε_t, for which convergence to zero follows immediately due to continuity of the parameterization. The other part can be bounded by the estimation error from estimating sample covariances of stationary processes. This expression can be bounded uniformly in the lag (see Hannan and Deistler, 1988, Theorem 5.3.2). The same decomposition as just mentioned can also be applied to the terms appearing in the other items. Consider e.g. $l(\overline{z_k},\tau)T^{-1}\sum_{t=1}^{T} x_t(z_k)\varepsilon_t' k(\overline{z_k},\tau)'$, which is the product of three terms. Of these three terms two are deterministic and depend continuously on the parameter vector, the third term is stochastic and independent of the parameter vector. This finishes the proof of the Lemma. □

The above lemmata provide the required technical results in order to prove Theorem 1. Consider the concentrated pseudo log likelihood:

$$L_T(\tau) = \log\det\left(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t(\tau,\hat{D})\varepsilon_t(\tau,\hat{D})'\right)$$

The log det transformation is monotonous, hence it suffices to investigate $T^{-1}\sum_{t=1}^{T}\varepsilon_t(\tau,\hat{D})\varepsilon_t(\tau,\hat{D})'$. Recall that $y_t = k^0(z)\varepsilon_t + D^0 s_t + C^0(A^0)^{t-1}x_1$, with variance of ε_t given by Σ^0.
Analogously to Saikkonen (1995) decompose

$$\begin{array}{rcl} \varepsilon_t(\tau,\hat{D}) & = & [k(z,\tau_u,\tau_{st})^{-1} - k(z,\tau_u^0,\tau_{st})^{-1}]k^0(z)\varepsilon_t + k(z,\tau_u,\tau_{st})^{-1}[C^0(A^0)^{t-1}x_1^0 + (D^0-\hat{D})s_t] \\ & & +\{k(z,\tau_u^0,\tau_{st})^{-1}[k(z,\tau_u^0,\tau_{st}^0) - k(z,\tau_u^0,\tau_{st})] + I\}\varepsilon_t \\ & = & \varepsilon_{t,u}(z,\tau_u,\tau_{st}) + \varepsilon_{t,st}(z,\tau_u^0,\tau_{st}) \end{array}$$

using again the separation of τ (and similarly of τ^0) in τ_u and τ_{st}. Note, that the second term, $\varepsilon_{t,st}(\tau_u^0,\tau_{st})$, does not depend on τ_u but only on the true value τ_u^0 and also note that $\varepsilon_{t,u}(\tau_u^0,\tau_{st}) =$

$k(z,\tau_u^0,\tau_{st})^{-1}[C^0(A^0)^{t-1}x_1^0+(D^0-\hat{D})s_t]$ for all values of τ_{st}. In the theorem two scenarios are given: Either $\hat{D}=D^0=0$ or $\hat{D}$ is included in the parameter vector. In the first case $\varepsilon_{t,u}(\tau_u^0,\tau_{st})=0$, whereas in the second case the contribution is nonzero. Hence we need to obtain an understanding concerning the asymptotic behavior of $\hat{D}(\tau,\Sigma)$ in this case. From the estimation equation (15) it follows that

$$k^{-1}(\overline{z_j},\tau)'\Sigma^{-1}\frac{1}{T}\sum_{t=1}^{T}(k^{-1}(z,\tau)(y_t-D^0s_t))\overline{s_{t,j}} = k^{-1}(\overline{z_j},\tau)'\Sigma^{-1}\sum_{r=1}^{m}\left(\frac{1}{T}\sum_{t=1}^{T}s_{t,r}\overline{s_{t,j}}\right)k^{-1}(\overline{z_r},\tau)(\hat{D}_r-D_r^0)$$

where $\hat{D}_r$ denotes the r-th column of $\hat{D}$ and D_r^0 the r-th column of D^0. The above equation holds for $j=1,\dots,m$, i.e. for all components of s_t. Define furthermore the column vector

$$\hat{\Gamma}(\tau,\Sigma)=\left[k^{-1}(\overline{z_j},\tau)'\Sigma^{-1}\frac{1}{T}\sum_{t=1}^{T}(k^{-1}(z,\tau)(y_t-D^0s_t))\overline{s_{t,j}}\right]_{j=1,\dots,m}$$

and the $ms\times ms$ matrix (j denotes the row index, r the column index)

$$\hat{\Delta}(\tau,\Sigma)=\left[k^{-1}(\overline{z_j},\tau)'\Sigma^{-1}k^{-1}(\overline{z_r},\tau)\frac{1}{T}\sum_{t=1}^{T}s_{t,r}\overline{s_{t,j}}\right]_{j,r=1,\dots,m}$$

Then consequently the following relationship holds:

$$\hat{\Gamma}(\tau,\Sigma)=\hat{\Delta}(\tau,\Sigma)\mathrm{vec}[\hat{D}-D^0] \tag{18}$$

From the definition of the quantities immediately $\hat{\Delta}(\tau,\Sigma)\geq 0$ follows. It also follows that it is only positive semidefinite, not positive definite. The estimate $\mathrm{vec}[\hat{D}-D^0]=\hat{\Delta}(\tau,\Sigma)^{\dagger}\hat{\Gamma}(\tau,\Sigma)$ is well defined, with $\dagger$ denoting the Moore-Penrose inverse, and a solution to the equation (18). Note that $T^{-1}\sum_{t=1}^{T}\overline{s_{t,j}}s_{t,r}\to 0$ for $j\neq r$ and that $y_t-D^0s_t=k^0(z)\varepsilon_t+C^0(A^0)^{t-1}x_1^0$. It also follows that $\hat{\Delta}(\tau,\Sigma)$ converges to $\mathrm{diag}[k^{-1}(\overline{z_j},\tau)'\Sigma^{-1}k^{-1}(\overline{z_j},\tau)]$, where convergence is uniform in the parameter space. Now decompose $\hat{\Sigma}(\tau)$ in the following way:

$$\begin{aligned}\hat{\Sigma}(\tau) &= \frac{1}{T}\sum_{t=1}^{T}(\varepsilon_{t,u}(\tau)\varepsilon_{t,u}(\tau)'+\varepsilon_{t,st}(\tau_u^0,\tau_{st})\varepsilon_{t,u}(\tau)'+\varepsilon_{t,u}(\tau)\varepsilon_{t,st}(\tau_u^0,\tau_{st})')+\\ &\quad+\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{t,st}(\tau_u^0,\tau_{st})\varepsilon_{t,st}(\tau_u^0,\tau_{st})'\\ &= Q_T^u(\tau)+Q_T^{st}(\tau_{st})\end{aligned}$$

The following lemma clarifies the properties of $\varepsilon_{t,st}(\tau_u^0,\tau_{st})$ and $Q_T^{st}(\tau_{st})$.

Lemma 4 *Under the assumptions of Theorem 1, $\varepsilon_{t,st}(\tau_u^0,\tau_{st})$ is asymptotically stationary for all $\tau_{st}\in\Theta_{st}^{\tau_u^0}$. If $\tilde{\tau}_{st}=\mathrm{argmin}_{\tau_{st}\in\Theta_{st}^{\tau_u^0}}\log\det Q_T^{st}(\tau_{st})$, then $\tilde{\tau}_{st}\to\tau_{st}^0$ a.s. Further, $\log\det Q_T^{st}(\tilde{\tau}_{st})=\log\det\Sigma^0+o(T^{-\alpha})$ for all $\alpha<1$. Therefore* $\inf\log\det Q_T^{st}(\tau_{st})\geq\log\det\Sigma^0+o(T^{-\alpha})$ *a.s.*

PROOF: Let $x_t^0=[(x_{t,u}^0)',(x_{t,st}^0)']'$, where $x_{t,u}^0\in\mathbb{C}^c$ denotes the nonstationary part of the state and $x_{t,st}^0\in\mathbb{C}^{n-c}$ denotes the stationary part of x_t. Perform the same decomposition in the stationary and nonstationary part also for $x_t(\tau)$. Further let $C=[C_u,C_{st}]$, where $C_u=[C_1,\dots,C_{l_{2\pi}}]$ is decomposed into the blocks of coordinates of the state corresponding to the different unit roots ordered according to increasing frequency. Note that $[k(z,\tau_u^0,\tau_{st}^0)-k(z,\tau_u^0,\tau_{st})]\varepsilon_t=C^0x_t^0-Cx_t(\tau)$ and therefore

$$\begin{aligned}\varepsilon_{t,st}(\tau_u^0,\tau_{st}) &= \varepsilon_t+k(z,\tau_u^0,\tau_{st})^{-1}[C^0x_t^0-Cx_t(\tau)]\\ &= \varepsilon_t+k(z,\tau_u^0,\tau_{st})^{-1}[C_u^0x_{t,u}^0-C_u^0x_{t,u}(\tau)+C_{st}^0x_{t,st}^0-C_{st}x_{t,st}(\tau)]\end{aligned}$$

Since $k(z,\tau_u^0,\tau_{st})^{-1}$ is uniformly stable due to the assumptions and ε_t, $x_{t,st}^0$ and $x_{t,st}(\tau)$ are stationary by definition, it follows that $\varepsilon_{t,st}(\tau_u^0,\tau_{st})$ is stationary, if $k(z,\tau_u^0,\tau_{st})^{-1}C_u^0(x_{t,u}^0 - x_{t,u}(\tau))$ is stationary. It is shown below that for $k=1,\ldots,l_{2\pi}$ one obtains $k(e^{i\omega_k},\tau_u^0,\tau_{st})^{-1}[C_k^0]=0$, proving that $\varepsilon_{t,st}(\tau_u^0,\tau_{st})$ is indeed stationary, since this zero cancels the integration in the corresponding block-component of x_t and $x_t(\tau)$. Therefore, standard theory for stationary processes applies in the estimation, proving the claims in the lemma.
In particular $\inf \log\det Q_T^{st}(\tau_{st}) \geq \log\det \Sigma^0 + o(T^{-\alpha})$ follows from the proof of consistency of the `BIC` order estimates (cf. Hannan and Deistler, 1988, Chapter 5). □

Now we are ready to commence the proof of super-consistency for the part of the parameter vector corresponding to C_u. The proof consists of two steps. In the first step (Lemma 5) consistency is established, and in the second step (Lemma 6) the result is sharpened to super-consistency. The proof of Theorem 1 is then concluded by showing consistency also for τ_{st} and investigating the behavior of $\hat{D}$.
As above, partition the matrix C_u according to the unit roots $C_u = [C_1,\ldots,C_{l_{2\pi}}]$. Let $N_{T,\gamma}(\tau_u^0,\delta) \subset \Theta_u$ denote the set of all matrices $C_u \in \Theta_u$, such that $\max_{k=1,\ldots,l_{2\pi}} \|(I - C_k C_k')C_k^0\| > \delta/T^\gamma$ with $0 \leq \gamma < 1$. Further, let $B_{\tau_u^0,\delta} = N_{T,0}(\tau_u^0,\delta)$. Then the following lemma establishes consistency of the estimate $\hat{C}_u$, i.e. the proof is performed independently of a parameterization:

Lemma 5 *Let the conditions of Theorem 1 hold. Then $dist(\hat{C}_u, C_u^0) \to 0$ in probability.*

`PROOF:` The infimum over all $\tau_u \in N_{T,\gamma}(\tau_u^0,\delta), \tau_{st} \in \Theta_{st}^{\tau_u}$ of $\log\det Q_T(\tau)$ can be bounded from below by $\inf_{\tau_u}\left[\log\det[Q_T^u(\tau)Q_T^{st}(\tau_{st})^{-1}+I] + \inf_{\tau_{st}\in\Theta_{st}^{\tau_u}} \log\det Q_T^{st}(\tau_{st})\right]$. Lemma 4 states that

$$\inf_{\tau_{st}\in\Theta_{st}^{\tau_u}} \log\det Q_T^{st}(\tau_{st}) = \log\det\Sigma^0 + o(T^{-\alpha}), \forall \alpha < 1$$

It follows from the compactness assumption on the parameter set, that $\sup_{\tau_{st}\in\Theta_{st}} \|Q_T^{st}(\tau_{st})\| < \infty$ a.s. Thus, it is sufficient to show that for all constants $M > 0$, some $0 \leq m$ and all $\delta > 0$

$$\lim_{T\to\infty} \mathbb{P}\left\{\inf \lambda_{max}(Q_T^u) \geq M\right\} = 1, \lim_{T\to\infty} \mathbb{P}\left\{\inf \lambda_{min}(Q_T^u) \geq m\right\} = 1 \tag{19}$$

where again the monotonicity property of the log det transformation is employed. Here λ_{max} and λ_{min} denote the largest and the smallest eigenvalue respectively.
Let $\langle a_t, b_t\rangle = T^{-1}\sum_{t=1}^T a_t b_t'$ for any sequences $(a_t)_{t\in\mathbb{N}}, (b_t)_{t\in\mathbb{N}}$, somewhat sloppily using the same symbol for the sequence and the element of the sequence. Here it is assumed that unavailable observations are taken to be zero. Then one obtains $Q_T^u = \langle \varepsilon_{t,u}(\tau), \varepsilon_{t,u}(\tau)\rangle + \langle \varepsilon_{t,u}(\tau), \varepsilon_{t,st}(\tau_u^0,\tau_{st})\rangle + \langle \varepsilon_{t,st}(\tau_u^0,\tau_{st}), \varepsilon_{t,u}(\tau)\rangle$, where

$$\varepsilon_{t,u}(\tau) = [k^{-1}(z,\tau_u,\tau_{st}) - k^{-1}(z,\tau_u^0,\tau_{st})][C_u^0(I - zJ_u^0)^{-1}B_u^0 z\varepsilon_t + k_{st}(z,\tau_{st}^0)\varepsilon_t] + k(z,\tau_u,\tau_{st})^{-1}[C^0(A^0)^{t-1}x_1^0 + (D^0 - \hat{D})s_t]$$

The matrix J_u^0 denotes as before the block of A^0 corresponding to the unit roots, where the blocks are in Jordan normal form and ordered according to increasing frequency in $[0,2\pi)$. Due to the assumed strict minimum-phase assumption the inverses all exist and are stable. Examine

$$k(z,\tau_u,\tau_{st})^{-1}(D^0 - \hat{D})s_t = \sum_{i=1}^m k(z,\tau_u,\tau_{st})^{-1}(D_i^0 - \hat{D}_i)s_{t,i} = \sum_{i=1}^m k(\overline{z_i},\tau_u,\tau_{st})^{-1}(D_i^0 - \hat{D}_i)s_{t,i}$$

more closely: From the definitions of the respective quantities it follows that

$$\begin{aligned}
\hat{\Delta}(\tau,\Sigma) &= \left[k^{-1}(\overline{z_j},\tau)'\Sigma^{-1}k^{-1}(\overline{z_r},\tau)\langle s_{t,r}, s_{t,j}\rangle\right]_{j,r=1,\cdots,m}, \\
\hat{\Gamma}(\tau,\Sigma) &= \left[k^{-1}(\overline{z_j},\tau)'\Sigma^{-1}\langle k^{-1}(z,\tau)(k^0(z)\varepsilon_t + C^0(A^0)^{t-1}x_1^0), s_{t,j}\rangle\right]_{j=1,\ldots,m}
\end{aligned}$$

where also $y_t - D^0 s_t = k^0(z)\varepsilon_t + C^0(A^0)^{t-1}x_1^0$ has been used.
Elementary calculations show, that $T^{-1}\langle \varepsilon_{t,u}(\tau), \varepsilon_{t,st}(\tau_u^0,\tau_{st})\rangle \to 0$ in probability uniformly on the

parameter space. Here the fact that $\varepsilon_{t,st}(\tau_u^0, \tau_{st})$ is stationary is used and the uniformity is shown using the results of Lemma 1 and 3. Hence the essential term is $T^{-1}\langle \varepsilon_{t,u}(\tau), \varepsilon_{t,u}(\tau)\rangle$. Since the term due to $k_{st}(z, \tau_{st}^0)\varepsilon_t$ is stationary, it is sufficient to focus on $[k^{-1}(z, \tau_u, \tau_{st}) - k^{-1}(z, \tau_u^0, \tau_{st})]C_u^0(I - zJ_u^0)^{-1}B_u^0 z\varepsilon_t = \sum_{k=1}^{l_{2\pi}} \tilde{k}_k(z, \tau)x_t^0(z_k)$, which defines the transfer functions $\tilde{k}_k(z, \tau) = [k^{-1}(z, \tau_u, \tau_{st}) - k^{-1}(z, \tau_u^0, \tau_{st})]C_k^0$. Here the variables $x_t^0(z_k)$ denote the components of the state $x_t^0(\tau)$ corresponding to the unit root z_k. Thus, $x_{t+1}^0(z_k) = z_k x_t^0(z_k) + B_k^0\varepsilon_t$, where $x_1^0(z_k) = 0$ for all k. Next examine the term $k^{-1}(z, \tau_u, \tau_{st})C^0(A^0)^{t-1}x_1^0 = \sum_{k=1}^{l_{2\pi}} k^{-1}(z, \tau_u, \tau_{st})C_k^0 z_k^{t-1}x_{1,k}^0 + k^{-1}(z, \tau_u, \tau_{st})C_{st}^0(A_{st}^0)^{t-1}x_{t,st}^0$, which is one component of $\varepsilon_{t,u}(\tau)$. Since $k^{-1}(\overline{z_k}, \tau_u^0, \tau_{st})C_k^0 = 0$ (see below), this term is equal to $\sum_{k=1}^{l_{2\pi}} \tilde{k}_k(z, \tau)z_k^{t-1}x_{1,k}^0 + k^{-1}(z, \tau_u, \tau_{st})C_{st}^0(A_{st}^0)^{t-1}x_{t,st}^0$. Hence with slight abuse of notation $x_t^0(z_k)$ is assumed to be started at $x_{1,k}^0$ from now on. Using an analogous argument, $k^{-1}(z, \tau)(k^0(z)\varepsilon_t + C^0(A^0)^{t-1}x_1^0)$ in the definition of $\hat{\Gamma}(\tau, \Sigma)$ can be replaced with $\sum_{k=1}^{l_{2\pi}} \tilde{k}_k(z, \tau)x_t^0(z_k)$. Hence the essential terms in $\varepsilon_{t,u}(\tau)$ is equal to

$$\sum_{k=1}^{l_{2\pi}} \tilde{k}_k(z, \tau)x_t^0(z_k) - [k(\overline{z_r}, \tau)^{-1}s_{t,r}][k^{-1}(\overline{z_j}, \tau)'\Sigma^{-1}k^{-1}(\overline{z_r}, \tau)\langle s_{t,r}, s_{t,j}\rangle]^{-1}[k^{-1}(\overline{z_j}, \tau)'\Sigma^{-1}\langle \tilde{k}_k(z, \tau)x_t^0(z_k), s_{t,j}\rangle] \tag{20}$$

where r runs over block-columns and j indicates the block-row index.
For notational convenience, let $w_t(z_k) = \tilde{k}_k(z, \tau)x_t^0(z_k)$ in the case, that $\hat{D} = D^0 = 0$ and $w_t(z_k)$ equal to the summand for fixed k in (20). Considering the order of magnitude implied by the various terms according to Lemma 3, it follows that the essential terms in $T^{-1}Q_T^u$ are

$$\begin{aligned} T^{-2}\sum_{t=1}^{T}\left[\sum_{k=1}^{l_{2\pi}} w_t(z_k)\right]\left[\sum_{k=1}^{l_{2\pi}} w_t(z_k)\right]' &= \sum_{k=1}^{l_{2\pi}}\left(T^{-2}\sum_{t=1}^{T} w_t(z_k)w_t(z_k)'\right) + o_P(1) \\ &\xrightarrow{d} \sum_{k=1}^{l_{2\pi}} \tilde{k}_k(\overline{z_k})W(z_k)\tilde{k}_k(\overline{z_k})' \end{aligned} \tag{21}$$

as the remaining terms are sample covariances of stationary processes or deterministic components with stationary or integrated processes and can thus be neglected for our purpose. The $o_P(1)$ indicates convergence to zero in probability and is due to the neglection of the cross terms, i.e. the terms $T^{-2}\sum_{t=1}^{T} w_t(z_k)w_t(z_j)'$ for $z_k \neq z_j$. Convergence of this term follows from Lemma 3 item iv) in the case of no included constant term and using items iv) and vi) else. In order to show weak convergence, consider $T^{-3/2}\sum_{t=1}^{T}\langle \tilde{k}_k(z, \tau)x_t^0(z_k), s_{t,j}\rangle \to 0$ for $k \neq j$ according to item vi) of Lemma 3. Hence only one block in $T^{-1/2}\hat{\Gamma}(\tau, \Sigma)$ has a nonzero limit. $\hat{\Delta}(\tau, \Sigma)$ converges to a block-diagonal matrix. The rest follows from straightforward but cumbersome calculations taking the singularity of $\hat{\Delta}(\tau, \Sigma)$ into account.
In the case of no included deterministic terms $W(z_k) = Z(z_k)$ as defined in Lemma 2, else $W(z_k) = Z(z_k) - Y(z_k)Y(z_k)'$, which can also be written as $\int_0^1 B_k(u)B_k(u)'du$, where $B_k(u)$ denotes the demeaned Brownian motion $B_k(u) = W_k(u) - \int_0^1 W_k(u)du$. Each of the terms of the limit expression in (21) is non-negative. So, to show the claim, it is sufficient to show that for at least one of the summands the probability that it is strictly positive tends to one. If that is established, the additional scaling with T^{-1} then implies that the largest eigenvalue of $Q_T^u(\tau)$ tends to infinity for $\tau_u \in B_{\tau_u^0,\delta}$. Thus, it suffices to show that $\inf_{\tau_u \in B_{\tau_u^0,\delta}} \max_{k=1,\ldots,l_{2\pi}} \|\tilde{k}_k(\overline{z_k}, \tau)\| > c > 0$. In order to do so, fix $\omega = \omega_k \in [0, 2\pi)$ for a moment. To simplify notation we rearrange the terms as follows. With subscript $_1$ we denote terms corresponding to the fixed unit root $z = e^{i\omega}$. With subscript $_2$ all other terms are denoted. For further simplification of the argument also the state (and the system matrix blocks) are reordered such that the components of $x_t(\omega)$ are the first components. Thus, consider

$$k^{-1}(e^{i\omega}, \tau_u, \tau_{st}) = I - [C_1, C_2]\begin{bmatrix} B_1C_1 & B_1C_2 \\ B_2C_1 & e^{-i\omega}I - A_2 + B_2C_2 \end{bmatrix}^{-1}\begin{bmatrix} B_1 \\ B_2 \end{bmatrix}$$

$$\begin{aligned} &= I - [C_1, C_2]\left\{\begin{bmatrix} 0 & 0 \\ 0 & \bar{A}_2^{-1} \end{bmatrix} + \begin{bmatrix} I \\ -\bar{A}_2^{-1}B_2C_1 \end{bmatrix} U^{-1}[I, -B_1C_2\bar{A}_2^{-1}]\right\}\begin{bmatrix} B_1 \\ B_2 \end{bmatrix} \\ &= k_2(e^{i\omega}) - k_2(e^{i\omega})C_1(B_1k_2(e^{i\omega})C_1)^{-1}B_1k_2(e^{i\omega}) \end{aligned}$$

where the dependence of all terms on τ is suppressed for notational simplicity. Here $\bar{A}_2 = e^{-i\omega}I - A_2 + B_2C_2, k_2(e^{i\omega}) = I - C_2\bar{A}_2^{-1}B_2$ and $U = (B_1C_1 - B_1C_2\bar{A}_2{}^{-1}B_2C_1) = B_1k_2(e^{i\omega})C_1$. Note that we assume that $k_2(e^{i\omega})$ exists, i.e. that $\bar{A}_2$ is invertible. This corresponds to the assumption that $k_2(z,\tau)$ has no pole at $e^{-i\omega}$ or equivalently that $k_2^{-1}(z,\tau)$ has no zero there. $k^{-1}(e^{i\omega}, \tau_u, \tau_{st})C_1 = 0$ and therefore we obtain $[k^{-1}(e^{i\omega}, \tau_u, \tau_{st}) - k^{-1}(e^{i\omega}, \tau_u^0, \tau_{st})]C_1^0 = k^{-1}(e^{i\omega}, \tau_u, \tau_{st})[C_1^0 - C_1]$. Thus, we obtain under the assumption on $k_2(z,\tau)$ that

$$(I - C_1C_1')k_2^{-1}(e^{i\omega})k^{-1}(e^{i\omega}, \tau_u, \tau_{st})[C_1^0 - C_1] = (I - C_1C_1')C_1^0$$

Note that in general $(I - C_1C_1') \neq 0$, as in a minimal representation there cannot be more than s common cycles for each unit root. In the boundary case of s stochastic cycles for a unit root ω, it follows due to the orthonormality and p.l.t. restrictions that $C_1 = I_s$. In this case $C_1 = C_1^0 = I_s$, as in this case no parameter has to be estimated in C_1.
The above result rests upon the assumption of a uniformly bounded $k_2^{-1}(z,\tau)$ at the unit roots $z = e^{i\omega}$. This is the reason for stating this assumption in the formulation of Theorem 1. Given the assumption is fulfilled, it follows that for $\tau_u \in B_{\tau_u^0,\delta}$ it holds that $\max_{k=1,\dots,l_{2\pi}} \|\tilde{k}_k(\bar{z}_k,\tau)\| \geq \max_{k=1,\dots,l_{2\pi}} c\|(I - C_kC_k')C_k^0\| \geq c\delta$. This shows that $\mathbb{P}(\hat{\tau}_u \in B_{\tau_u^0,\delta}) \to 0$ for arbitrary $\delta > 0$. Hence, weak consistency of τ_u is established.□

In the next step we establish super-consistency of τ_u, the parameters corresponding to C_u.

Lemma 6 *Under the conditions of Theorem 1, $T^\gamma dist(\hat{C}_u - C_u^0) \to 0$ in probability for all $0 < \gamma < 1$.*

PROOF: The proof resembles the consistency proof given in Lemma 5 and differs essentially only in the derivation of the bound for $\tilde{k}_k(\overline{z_k},\tau)$. Similar to the consistency proof it is sufficient to show that

$$\lim_{T\to\infty} \mathbb{P}\{\inf \lambda_{max}(T^\alpha Q_T^u) \geq M\} = 1, \lim_{T\to\infty} \mathbb{P}\{\inf \lambda_{min}(Q_T^u) \geq m\} = 1 \tag{22}$$

for all $0 < \alpha < 1$, where the infimum is taken over the set $\overline{N_{T,\gamma}(\tau_u^0,\delta)}$, which renders $\tilde{k}_k(\overline{z_k},\tau) \to 0$ possible. Consider a sequence $\tau_{u,T} \in N_{T,\gamma}(\tau_u^0,\delta)$ and let $c_T(\gamma) = \max_{k=1,\dots,l_{2\pi}} \|(I - C_k(\tau_{u,T})'C_k(\tau_{u,T}))C_k^0\|$, where the dependence of $C_k(\tau_{u,T})$ on the parameter vector $\tau_{u,T}$ is emphasized. Note that $c_T(\gamma) \geq \delta T^{-\gamma}$ due to the definition of $N_{T,\gamma}(\tau_u^0,\delta)$. From the expressions given above it follows that

$$\begin{aligned} \limsup_{T\to\infty} c_T(\gamma)^{-1} \max_{k=1,\dots,l_{2\pi}} \|\tilde{k}_k(\overline{z_k}, \tau_{u,T}, \tau_{st,T})\| &< \infty \\ \liminf_{T\to\infty} c_T(\gamma)^{-1} \max_{k=1,\dots,l_{2\pi}} \|\tilde{k}_k(\overline{z_k}, \tau_{u,T}, \tau_{st,T})\| &> c > 0 \end{aligned}$$

uniformly in $\tau_{st} \in \Theta_{st}$ due to the assumptions on the transfer functions. From Lemma 1 and Lemma 2 we obtain also

$$\mathbb{P}\left\{\|\frac{1}{T^2c_T^2(\gamma)}\sum_{t=1}^T w_t(z_k)\left([k^{-1}(z,\tau_{u,T},\tau_{st,T}) - k^{-1}(z,\tau_u^0,\tau_{st,T})]k_{st}(z,\tau_{st,T})\varepsilon_t + \varepsilon_{t,st}(\tau_{st})\right)'\| > 0\right\} \to 0 \tag{23}$$

where $w_t(z_k)$ is defined in the proof of Lemma 5. In case that the deterministic variables are not included in y_t and the estimation, i.e. $\hat{D} = D^0 = 0$, this follows from $\tilde{k}_k(\overline{z_k}, \tau_{u,T}, \tau_{st,T})/c_T(\gamma) < \infty$ and the convergence results on sample moments between processes integrated of order 1 and stationary processes. When the deterministic terms are included in the estimation, the arguments are more cumbersome, but still standard and thus omitted. In this expression the fact $1/(Tc_T(\gamma)) = o(1)$ due to $\gamma < 1$ is used. This implies that the squared terms dominate and that the essential term in $T^\alpha Q_T^u$ is equal to

$$
\begin{aligned}
&T^{\alpha-1}\sum_{t=1}^T \left(\sum_{k=1}^{l_{2\pi}} w_t(z_k)\right)\left(\sum_{k=1}^{l_{2\pi}} w_t(z_k)\right)' \geq \frac{T^\epsilon}{c_T^2(\gamma)T^2}\sum_{t=1}^T \left(\sum_{k=1}^{l_{2\pi}} w_t(z_k)\right)\left(\sum_{k=1}^{l_{2\pi}} w_t(z_k)\right)' \\
&\geq \frac{T^\epsilon}{c_T^2(\gamma)} \left[\begin{array}{ccc} \tilde{k}_1(\bar{z}_1) & \cdots & \tilde{k}_{l_{2\pi}}(\bar{z}_{l_{2\pi}}) \end{array}\right] \left(\left[\begin{array}{cccc} W(z_1) & 0 & \ldots & 0 \\ 0 & W(z_2) & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & W(z_{l_{2\pi}}) \end{array}\right] + o_P(1) \right) \left[\begin{array}{c} \tilde{k}_1(\bar{z}_1)' \\ \vdots \\ \tilde{k}_{l_{2\pi}}(\bar{z}_{l_{2\pi}})' \end{array}\right] \\
&\geq T^\epsilon \left(\max_{k=1,\ldots,l_{2\pi}} \|\tilde{k}_k(\bar{z}_k)\| c_T(\gamma)^{-1}\right)^2 (\lambda_{min}(X) + o_P(1))\, I_s
\end{aligned}
$$

for some ϵ such that $0 < \epsilon \leq \alpha - 1 + 2(1-\gamma)$ and X denotes the matrix in brackets in the center of the expression in the line above. For notational convenience the dependence of $\tilde{k}_k(\bar{z}_k)$ on τ is neglected. The results of Johansen and Schaumburg (1999) imply that the probability of the smallest eigenvalue of this matrix X to be positive tends to one, whether or not deterministic terms are included in the estimation. This shows that

$$\mathbb{P}\left\{ \inf_{\tau_u \in N_{T,\gamma}(\tau^0,\delta), \tau_{st}\in\Theta_{st}} |\lambda_{max}(T^\alpha Q_T^u(\tau_u,\tau_{st}))| > M \right\} \to 1,$$

since the result has been shown for all parameter values in the closure of the notified set. Because $0 < \alpha < 1$ can be chosen arbitrarily, this proves consistency of order T^γ for $0 < \gamma < 1$.
Note again that the proof relies solely upon the matrix C_u itself, at no point the parameter vector τ_u has been employed. This amounts to say that the result is parameterization independent.□

Now, in the next to last step the asymptotic behavior of $\hat{D}$ is analyzed:

Lemma 7 *Partition the components of $s_t = [(s_t^1)', (s_t^2)']' \in \mathbb{C}^m$, where again in s_t^1 the cyclical components $s_{t,j}^1$ for $j = 1, \ldots, l_{2\pi}$ corresponding to the unit root frequencies are collected and in s_t^2 (components $l_{2\pi}+1, \ldots, m$ of s_t the cyclical components to the non unit root frequencies are collected. Then for $j = 1, \ldots, l_{2\pi}$ denote with $\Pi_j = C_j^0(C_j^0)'$ for $j = 1, \ldots, l_{2\pi}$ and $\Pi_j = 0^{s\times s}$ for $j = l_{2\pi}+1, \ldots, m$. Then it holds for all $j = 1, \ldots, m$ that*

$$\Pi_j \hat{D}_j \to 0$$

$$(I - \Pi_j)\hat{D}_j \to (I-\Pi_j)D_j^0$$

where D_j^0 denotes again the j-th column of D^0.

PROOF: We have already established that $\hat{\Sigma}$ is bounded and has bounded inverse almost surely. It has also been shown that $\hat{\Delta}(\tau,\Sigma)$ converges to $\text{diag}[k^{-1}(\overline{z_j},\tau)'\Sigma^{-1}k^{-1}(\overline{z_j},\tau)]$. Therefore each component can be analyzed separately.
From the established super-consistency for $\hat{C}_u$ for the j-th block-entry of $\hat{\Gamma}(\tau,\Sigma) \to 0$ follows, as it has been shown to be equal to

$$k^{-1}(\overline{z_j},\tau)'\Sigma^{-1}\sum_{s=1}^{l_{2\pi}} \tilde{k}_s(\overline{z_s},\tau)\frac{1}{T}\sum_{t=1}^T x_t^0(z_s)\overline{s_{t,j}}$$

where $\|\tilde{k}_s(\overline{z_s},\tau)\| = o_P(T^{-1/2})$ for $\gamma > 1/2$. This shows that $[k^{-1}(\overline{z_j},\tau)'\Sigma^{-1}k^{-1}(\overline{z_j},\tau)](\hat{D}_j - D_j^0) = k^{-1}(\overline{z_j},\tau_u^0,\tau_{st})'\Sigma^{-1}k^{-1}(\overline{z_j},\tau_u^0,\tau_{st})(\hat{D}_j - D_j^0) + o_P(1) \to 0$ uniformly in τ_{st}. The result now follows from the fact that the kernel of $k^{-1}(\overline{z_j},\tau_u^0,\tau_{st})$ is spanned by C_j^0 independently of τ_{st}. This holds for $j = 1, \ldots, l_{2\pi}$.
For the deterministic terms not corresponding to unit roots, the matrix $k^{-1}(\overline{z_j},\tau_u^0,\tau_{st})$ is nonsingular and therefore these columns can be estimated consistently. This concludes the proof of the lemma.□

Remark 6 *Any nonzero components of D_k^0 in the space spanned by the columns of C_k^0 cannot be expected to be estimated consistently due to the nonidentifiability involving $C_k^0 z_k^{t-1} x_1^0(z_k)$. However, by restricting these components to zero, which can be done without restriction of generality (see the main part of the paper), we achieve consistency for the matrix $\hat{D}$.*

The final step is now proving consistency of τ_{st}. From the results presented in the above lemmata, it follows that only $\tau_u \in \overline{N_{T,\gamma}(\tau_u^0, \delta)}$ have to be considered, where the *bar* here denotes the complement. Consider again the quantity Q_T^u. The above equations can be used in connection with Lemma 1 and Lemma 2 to show uniform convergence of Q_T^u to zero on the set $\overline{N_{T,\gamma}(\tau_u^0, \delta)}$ under the assumption $\gamma > 1/2$. Under this assumption it holds that $\|\tilde{k}_k(\bar{z}_k, \tau)\| \leq cT^{-\gamma}$. The evaluations are straightforward and thus omitted. This implies that the estimate $\hat{\tau}_{st}$ is essentially only a function of Q_T^{st} and does not depend upon τ_u and D. The problem can thus be analyzed by applying Lemma 4, in which consistency is derived for exactly the problem at hand. This finally concludes the proof of Theorem 1.

Proof of Theorem 2 (Asymptotic Distribution)

Let us repeat also at this point, that in order to derive the asymptotic distribution, the results are based on using a specific parameterization, whereas the consistency proof has been parameterization free. This approach allows us to embed the problem in some $\mathbb{R}^{d_1}$, for an appropriate d_1. Again we will use complex matrices in the proof, although we are concerned only with real valued processes. The relationships between complex and real valued representation have been discussed in detail in the main text.

For the applicability of linearization techniques we furthermore assume that the parameters are introduced in such a way that the true parameter vector τ^0 correspond to an interior point of $\Theta \subset \mathbb{R}^{d_1}$, with Θ an open subset. Due to the already established consistency it follows that for T large enough, the probability that the estimate $\hat{\tau}$ is contained in Θ tends to 1.

It is a well known fact in many situations that the asymptotic distribution of parameter estimates can depend upon the inclusion or exclusion of deterministic variables in the estimation. It will be seen below that this is also the case here. Collect all parameters together in $\alpha = (\tau', d', \sigma')'$ where $\tau = [\tau_u', \tau_{st}']'$, d is the parameter vector corresponding to D and $\sigma = \text{vech}(\Sigma)$, where $\text{vech}(X)$ denotes the operator stacking the diagonal and sub-diagonal elements of a symmetric matrix. $\sigma \in \mathbb{R}^{s(s+1)/2}$ also has to fulfill the restriction that the resulting Σ is positive definite, this restriction is fulfilled on an open subset of the Euclidean space and is of no further concern due to consistency. Consequently we denote by $\hat{\alpha} = (\hat{\tau}', \hat{d}', \hat{\sigma}')'$ the pseudo maximum likelihood estimate of α, i.e. the minimizer of $L_T(\alpha)$ and by $\alpha^0 = ((\tau^0)', (d^0)', (\sigma^0)')'$ the true parameter vector.

Now since τ^0 is assumed to be an interior point of Θ, it follows that for T large enough $\hat{\tau} \to \tau^0$ is an interior point as well. This implies that also $\hat{\alpha} \to \alpha^0$ is an interior point. Thus, a necessary condition for an optimum is a zero first derivative at the optimum:

$$\partial L_T(\hat{\alpha}) = 0 = \partial L_T(\alpha^0) + \partial^2 L_T(\bar{\alpha}_T)[\hat{\alpha} - \alpha^0],$$

where $\bar{\alpha}_T$ denotes an intermediate point between $\hat{\alpha}$ and α^0, not necessarily the same in each row. Let $D_T = \text{diag}(I, T^{1/2}I)$, where the sizes of the two blocks of D_T are equal to the dimensions of the parameter vectors τ_u and $[\tau_{st}', d', \sigma']'$ respectively. Further let $\tilde{D}_T = \text{diag}(TI, T^{1/2}I)$. Then the proof of the theorem proceeds in three steps:

1. Show that $D_T \partial L_T(\alpha^0)$ converges in distribution
2. Show that $D_T \partial^2 L_T(\bar{\alpha}_T) \tilde{D}_T^{-1}$ converges in distribution to a random matrix Z
3. Show that $\mathbb{P}\{Z > 0\} = 1$

Let us start with the first item, i.e. with establishing the asymptotic properties of the score vector. Denote with $\partial_i f(\alpha^0)$ the partial derivative of a function f with respect to the i-th component of the parameter vector α, evaluated at the point $\alpha = \alpha^0$. With subscript $i = st$ we denote the vector

of $\partial_i f(\alpha^0)$ for all i, such that the component τ_i is contained in τ_{st}. With subscript D we denote differentiation with respect to the entries in d, which parameterize the matrix D.
The asymptotic distribution of $\hat{\sigma}$ is not required in establishing the asymptotic distribution of the other parameters and is also not of any other interest to us. In the following lemma the asymptotic behavior of the score is summarized, ignoring the derivatives with respect to the entries in σ.

Lemma 8 *Let the assumptions of Theorem 2 hold. Then the following statements hold true:*

- $\sqrt{T}\partial_{st}L_T(\alpha^0) \xrightarrow{d} \mathcal{N}(0, V_{st})$, *where V_{st} denotes the asymptotic variance matrix and $\xrightarrow{d}$ denotes convergence in distribution.*
- *If i is such that τ_i corresponds to $C_k(\tau)$, then*

$$\partial_i L_T(\alpha^0) \xrightarrow{d} 2/\delta_k^2 \mathcal{R}\left\{ tr[(\Sigma^0)^{-1}\left(dk_i(\bar{z}_k, \tau^0)X(\omega_k)\right)]\right\}$$

where $dk_i(z,\tau) = -\partial_i C_k^0 - C(I - z(A^0 - B^0C^0))^{-1}B^0 z\partial_i C_k = -k^{-1}(z,\tau^0)\partial_i C_k^0$, with $X(\omega_k)$ as defined in Lemma 2.

- $\sqrt{T}\partial_D L_T(\alpha^0) \xrightarrow{d} \mathcal{N}(0, V_D)$. *For a column D_k of D corresponding to $z_k = \pm 1$, one obtains*

$$\sqrt{T}\partial_{D_k}L_T(\alpha^0) \xrightarrow{d} -2(k^0(\overline{z_k})^{-1})'(\Sigma^0)^{-1}W_k(1)$$

If $z_k \neq \pm 1$, the vector of derivatives with respect to the real part of the k-th column of D and with respect to the imaginary part of the same column converges to

$$\left[\begin{array}{c} -4\mathcal{R}\left\{(k^0(\overline{z_k},\tau^0)^{-1})'(\Sigma^0)^{-1}W_k(1)\right\} \\ -4\mathcal{I}\left\{(k^0(\overline{z_k},\tau^0)^{-1})'(\Sigma^0)^{-1}W_k(1)\right\} \end{array}\right]$$

where $W_k(1) = W_k^r(1) + iW_k^i(1)$ is as defined in Lemma 2. Hence the matrix V_D is block diagonal, where the diagonal blocks are given by the covariance matrices of the two parts of the vector given above, taking into account the uncorrelatedness of $W_k^r(1)$ and $W_k^i(1)$.

- *All convergence results hold jointly.*

`PROOF`: In order to establish the asymptotic properties of the score, the partial derivatives of $\varepsilon_t(\alpha)$ are required. These can be derived from the system equations:

$$\partial_i L_T(\alpha^0) = \partial_i \log\det\Sigma^0 - \mathrm{tr}[(\Sigma^0)^{-1}(\partial_i\Sigma^0)(\Sigma^0)^{-1}\frac{1}{T}\sum_{t=1}^T \varepsilon_t\varepsilon_t'] + \mathrm{tr}[(\Sigma^0)^{-1}\frac{2}{T}\sum_{t=1}^T \partial_i\varepsilon_t(\alpha^0)\varepsilon_t(\alpha^0)']$$

Here and also below the matrices (A, B, C) and the state x_t correspond to the real valued canonical representation of the true system described in Section 3. Recall that the matrix $C = [C_1, \ldots, C_l, C_{st}]$ is partitioned according to the blocks of different unit roots (or pairs of complex roots). Hence $C_k(z_k) \in \mathbb{R}^{s\times c_k(z_k)}$. Note that we here use the notation $C_k(z_k)$ to denote the real valued matrix blocks, whereas in the text this notation was only used for the complex canonical representation and no specific label was given to the real matrices. Accordingly $x_t(z_k)$ denotes the corresponding components of the state vector. Note that in the real representation:

$$x_t(z_k) = \left[\begin{array}{c} 2\mathcal{R}(z_k^{t-1}x_{1,k} + \sum_{l=1}^{t-1} z_k^l B_k\varepsilon_{t-l}) \\ 2\mathcal{I}(z_k^{t-1}x_{1,k} + \sum_{l=1}^{t-1} z_k^l B_k\varepsilon_{t-l}) \end{array}\right]$$

Let us start with the coordinates of $\tau_{st} = [\tau_B', \bar{\tau}_{st}']'$ and specifically with the parameters in $\bar{\tau}_{st}$. For these we obtain:

$$\partial_i A - B\partial_i C = \left[\begin{array}{cc} 0 & -B_u\partial_i C_{st} \\ 0 & \partial_i A_{st} - B_{st}\partial_i C_{st} \end{array}\right], \partial_i B = \left[\begin{array}{c} 0 \\ \partial_i B_{st} \end{array}\right], \partial_i C = \left[\begin{array}{cc} 0 & \partial_i C_{st} \end{array}\right]$$

The above equations together with the fact that $x_{t,st}(\tau^0)$ is stationary, imply stationarity of $\partial_i \varepsilon_t(\alpha)$ for all partial derivatives corresponding to entries in $\bar{\tau}_{st}$. With respect to the parameters collected in τ_B, which parameterize B_u, stationarity of the corresponding score components follows from $\partial_i A = 0$ and $\partial_i C = 0$. Now after having established stationarity for the components of the score vector corresponding to τ_{st}, asymptotic normality follows from well established theory for stationary processes, see e.g. Hannan and Deistler (1988). It is straightforward to show that the result holds jointly in all coordinates of τ_{st}.

Let us next analyze τ_u, i.e. the parameters attached to the matrices $C_k(z_k)$. The partial derivatives are:

$$\begin{aligned}\partial_l \varepsilon(\alpha) &= -(\partial_l C_k(z_k))x_{t,k} - C(\partial_l x_t(\tau)) \\ \partial_l x_{t+1}(\tau) &= -B(\partial_l C_k(z_k))x_{t,k} + (A - BC)\partial_l x_t\end{aligned}$$

These components of the score are filtered version of $x_{t,k}$. A possibly non-minimal representation of the filter is given by $dk_l(z,\tau) = -\partial_l C_k(z_k) - C(I - z(A-BC))^{-1}Bz\partial_l C_k(z_k)$, which is real valued. This representation allows for the application of Lemma 2 and Lemma 3 to obtain

$$\partial_l L_T(\alpha^0) = \operatorname{tr}[\Sigma^{-1}\frac{2}{T}\sum_{t=1}^{T}\partial_l \varepsilon_t(\alpha^0)\varepsilon_t'] + o_P(1) \xrightarrow{d} 2\mathcal{R}\left\{\operatorname{tr}\left((\Sigma^0)^{-1}dk_l(\overline{z_k},\tau^0)B_k\int_0^1 W_k(dW_k)')\right)\right\}$$

Now only the asymptotic distribution of the score components corresponding to d is left to be derived. The matrix D is parameterized with real parameters using $D = D_r + iD_i$, where both D_r and D_i are unconstrained, except for the restriction, that the columns of D_i corresponding to $z_k = \pm 1$ are zero. Consider a specific element of this part of the score vector, corresponding to component i, say, of d, which corresponds to entry (a,k) in D_r. Because of the restriction to real valued output processes, only the real part of the derivative has to be investigated:

$$\partial_i L_T(\alpha^0) = -\frac{2}{T}\sum_{t=1}^{T}(k^{-1}(z,\tau^0)e_a s_{t,k})'(\Sigma^0)^{-1}\varepsilon_t + o_P(T^{-1/2})$$

where e_a denotes the a-th vector of the canonical basis and $s_{t,k} = z_k^t$ is the k-th coordinate of s_t. It follows from the definition of $s_{t,k}$ that $k^{-1}(z,\tau^0)e_a s_{t,k} = k^{-1}(\overline{z_k},\tau^0)e_a z_k^{t-1}$. Therefore it follows that for $z_k = \pm 1$

$$\sqrt{T}\partial_i L_T(\alpha^0) = -e_a' k^{-1}(\overline{z_k},\tau^0)'(\Sigma^0)^{-1}\frac{2}{\sqrt{T}}\sum_{t=1}^{T} z_k^t \varepsilon_t \xrightarrow{d} -2e_a' k^{-1}(\overline{z_k},\tau^0)'(\Sigma^0)^{-1}W_k(1)$$

For $z_k \neq \pm 1$ one obtains

$$\sqrt{T}\partial_i L_T(\alpha^0) = -e_a'\mathcal{R}\left\{k^{-1}(\overline{z_k},\tau^0)'(\Sigma^0)^{-1}\frac{2}{\sqrt{T}}\sum_{t=1}^{T}\overline{z_k}^t\varepsilon_t)\right\} \xrightarrow{d} -4\delta_k e_a'\mathcal{R}\left\{k^{-1}(\overline{z_k},\tau^0)'(\Sigma^0)^{-1}W_k(1)\right\}$$

Note finally that if the derivative is with respect to the (a,k)-th entry in D_i, in the above equation $\mathcal{R}$ has to be replaced with $\mathcal{I}$. This concludes the proof of the lemma. □

After having established the (asymptotic) properties of the score vector, the next step is the analysis of the asymptotic behavior of the Hessian. Again the system equations can be used to obtain the following expressions for the second order partial derivatives:

$$\begin{aligned}\partial_{i,j}^2 \varepsilon_t(\alpha) &= -(\partial_{i,j}^2 C)x_t(\tau) - \partial_i C\partial_j x_t(\tau) - \partial_j C\partial_i x_t(\tau) - C\partial_{i,j}^2 x_t(\tau) \\ \partial_{i,j}^2 x_{t+1}(\tau) &= (\partial_{i,j}^2(A-BC))x_t(\tau) + \partial_i(A-BC)\partial_j x_t(\tau) + \partial_j(A-BC)\partial_i x_t(\tau) \\ &\quad +(A-BC)\partial_{i,j}^2 x_t(\tau) + \partial_{i,j}^2 By_t\end{aligned}$$

As in Lemma 8 in the discussion we have to distinguish with respect to which parameter components τ_u, $\bar{\tau}_{st}$ or d differentiation takes place. In addition to the previous lemma, we also have to consider the cross terms, where differentiation takes place e.g. once with respect to an entry in τ_u and once with respect to an entry in $\bar{\tau}_{st}$.

Lemma 9 *Under the conditions of Theorem 2 one obtains* $D_T \partial^2 L_T(\bar{\alpha}_T)\tilde{D}_T^{-1} \xrightarrow{d} Z$ *for each sequence* $\bar{\alpha}_T \to \alpha^0$.
In case that no deterministic terms are included in the true process and the estimation (i.e. $\hat{D} = D^0 = 0$*),* $Z = diag(Z_u, Z_{st}, Z_\Sigma)$ *is block diagonal. It holds that* $Z_{st} > 0$ *and* Z_Σ *are constant matrices and* Z_u *a random matrix, for which* $\mathbb{P}\{Z_u > 0\} = 1$ *holds.*
If the deterministic terms are included in the estimation, the following asymptotic distribution is obtained: Here again $\alpha = [\tau_u', \tau_{st}', d', \sigma']'$*. Then*

$$D_T \partial^2 L_T(\alpha_T)\tilde{D}_T^{-1} \xrightarrow{d} \begin{bmatrix} Z_u & 0 & Y_D' & 0 \\ 0 & Z_{st} & 0 & 0 \\ Y_D & 0 & Z_D & 0 \\ 0 & 0 & 0 & Z_\Sigma \end{bmatrix}$$

Z_u, Z_D *and* Y_D *are block-diagonal, with the diagonal blocks corresponding to different unit roots. For typical indices* i, j *(not the same in the expressions below) corresponding to the same unit root* $z_k = \pm 1$ *the respective entries are of the form:*

$$\begin{aligned} [Z_u]_{i,j} &= 2tr\left[\partial_j C_k'(k^0(\overline{z_k},\tau^0)^{-1})'(\Sigma^0)^{-1}k^0(\overline{z_k},\tau^0)^{-1}\partial_i C_k Z(z_k)\right] \\ [Z_D]_{i,j} &= 2e_i'(k^0(\overline{z_k},\tau^0)^{-1})'(\Sigma^0)^{-1}k^0(\overline{z_k},\tau^0)^{-1}e_j \\ [Y_D]_{i,j} &= -2e_i'(k^0(\overline{z_k},\tau^0)^{-1})'(\Sigma^0)^{-1}k^0(\overline{z_k},\tau^0)^{-1}\partial_j C_k Y(z_k) \end{aligned}$$

For the entries corresponding to $z_k \neq \pm 1$ *the respective expressions are:*

$$\begin{aligned} [Z_u]_{i,j} &= 4tr\mathcal{R}\left[\partial_j C_k'(k^0(\overline{z_k},\tau^0)^{-1})'(\Sigma^0)^{-1}k^0(\overline{z_k},\tau^0)^{-1}\partial_i C_k Z(z_k)\right] \\ [Z_D]_{i,j} &= 4\begin{bmatrix} e_i'(k^0(\overline{z_k},\tau^0)^{-1})'(\Sigma^0)^{-1}k^0(\overline{z_k},\tau^0)^{-1}e_j & 0 \\ 0 & e_i'(k^0(\overline{z_k},\tau^0)^{-1})'(\Sigma^0)^{-1}k^0(\overline{z_k},\tau^0)^{-1}e_j \end{bmatrix} \\ [Y_D]_{i,j} &= -4\begin{bmatrix} e_i'\mathcal{R}\left\{(k^0(\overline{z_k},\tau^0)^{-1})'(\Sigma^0)^{-1}k^0(\overline{z_k},\tau^0)^{-1}\partial_j C_k Y(z_k)\right\} \\ e_i'\mathcal{I}\left\{(k^0(\overline{z_k},\tau^0)^{-1})'(\Sigma^0)^{-1}k^0(\overline{z_k},\tau^0)^{-1}\partial_j C_k Y(z_k)\right\} \end{bmatrix} \end{aligned}$$

where for the entries in D *the respective entry in* D_r *and in* D_i *are given.*
It follows that $Z_u - Y_D' Z_D^\dagger Y_D$ *has the same structure as* Z_u*, where in the expression* $Z(z_k)$ *has to be replaced by* $Z(z_k) - Y(z_k)Y(z_k)'$*. Further* $Z_{st} > 0$ *and* $\mathbb{P}\{Z_u > 0\} \to 1$ *respectively* $\mathbb{P}\{Z_u - Y_D' Z_D^\dagger Y_D > 0\} \to 1$.

PROOF: In the proof first convergence of the various parts is shown and in a final step the nonsingularity of Z_u is established. First note that:

$$\begin{aligned} &\partial_{i,j}^2 L_T(\alpha) = \partial_{i,j}^2 \log\det\Sigma + \mathrm{tr}[\partial_{i,j}^2(\Sigma^{-1})\tfrac{1}{T}\textstyle\sum_{t=1}^T \varepsilon_t(\alpha)\varepsilon_t(\alpha)'] + \mathrm{tr}[\partial_i\Sigma^{-1}\tfrac{2}{T}\textstyle\sum_{t=1}^T \partial_j\varepsilon_t(\alpha)\varepsilon_t(\alpha)'] + \\ &\mathrm{tr}[\partial_j\Sigma^{-1}\tfrac{2}{T}\textstyle\sum_{t=1}^T(\partial_i\varepsilon_t(\alpha))\varepsilon_t(\alpha)'] + \mathrm{tr}[\Sigma^{-1}\tfrac{2}{T}\textstyle\sum_{t=1}^T(\partial_{i,j}^2\varepsilon_t(\alpha))\varepsilon_t(\alpha)'] + \mathrm{tr}[\Sigma^{-1}\tfrac{2}{T}\textstyle\sum_{t=1}^T(\partial_i\varepsilon_t(\alpha))(\partial_j\varepsilon_t(\alpha))'] \end{aligned} \tag{24}$$

According to the partitioning of α in four sub-vectors in total ten matrix blocks (taking into account symmetry of the Hessian) have to be dealt with. The blocks are partitioned according to how often differentiation takes place with respect to a component of τ_u, τ_{st}, d and σ.
The multiplication of the Hessian with D_T and $\tilde{D}_T$ has the following effect: For each derivative with respect to an entry in τ_u an additional scaling factor $T^{-1/2}$ is introduced, which results in the proper scaling factor for each of the terms.
In the above expression (24) the variable $\varepsilon_t(\alpha)$ appears, in the third, fourth and fifth term to be precise. This variable has to be evaluated at the point $\bar{\alpha}_T$. Due to the assumptions $\bar{\alpha}_T$ converges

to α^0. Hence, applying a mean value expansion again $\varepsilon_t(\bar{\alpha}_T) = \varepsilon_t + \partial\varepsilon_t(\tilde{\alpha})(\bar{\alpha}_T - \alpha^0)$, for suitable intermediate value $\tilde{\alpha}$, it follows that all three mentioned terms converge to 0. Look for example at the fifth term:

$$\frac{2}{T}\sum_{t=1}^{T}(\partial^2_{i,j}\varepsilon_t(\bar{\alpha}_T))\varepsilon_t(\bar{\alpha}_T)' = \frac{2}{T}\sum_{t=1}^{T}(\partial^2_{i,j}\varepsilon_t(\bar{\alpha}_T))\varepsilon_t' + \sum_{l=1}^{dim(\alpha)}\frac{2}{T}\sum_{t=1}^{T}(\partial^2_{i,j}\varepsilon_t(\bar{\alpha}_T))\partial_l\varepsilon_t(\tilde{\alpha})'(\bar{\alpha}_{l,T} - \alpha_l^0)$$

Lemmas 2 and 3 show that for this term for all possible combinations of differentiation (including the necessary normalization if differentiation occurs with respect to an entry of τ_u), that the first term of the above equation converges to 0. Due to the established *condition USE* this convergence is uniformly in a compact neighborhood of τ^0. Analogous considerations deliver convergence of the second term to 0 as well. Here the terms $(\partial^2_{i,j}\varepsilon_t(\alpha))\partial_l\varepsilon_t(\tilde{\alpha})'$ converge to random variables, post-multiplying with $(\bar{\alpha}_{l,T} - \alpha_l^0)$ then delivers the result. Similar considerations also apply to the third and fourth term of equation (24). Hence, we obtain:

$$\partial^2_{i,j}L_T(\alpha) = \partial^2_{i,j}\log\det\Sigma + \mathrm{tr}[\partial^2_{i,j}(\Sigma^{-1})\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t(\alpha)\varepsilon_t(\alpha)'] + \mathrm{tr}[\Sigma^{-1}\frac{2}{T}\sum_{t=1}^{T}(\partial_i\varepsilon_t(\alpha))(\partial_j\varepsilon_t(\alpha))'] + o_P(T^{N_u/2}) \tag{25}$$

where N_u counts the number of times differentiation takes place with respect to an element of τ_u. Now starting from equation (25) the asymptotic behavior of the derivatives can be analyzed.

Consider first differentiation with respect to entries of σ. If both i and j correspond to entries of σ, then it follows that $\partial_i\varepsilon_t(\alpha) = 0$. Thus, only the first two terms in (25) are relevant. These two converge to constants, noting that from the consistency proof we know $T^{-1}\sum_{t=1}^{T}\varepsilon_t(\bar{\alpha}_T)\varepsilon_t(\bar{\alpha}_T)' \to \Sigma^0$. For the case of first differentiating with respect to an entry of σ and then with respect to entry of α not part of σ the above equation directly delivers convergence to zero. This shows also that the estimates for Σ (or Σ^0 to stick with notation) and for the other parameters are asymptotically uncorrelated. Thus, the asymptotic distribution of $\hat{\Sigma}$ is indeed not of interest with respect to the asymptotic distribution of the remaining parameters.

If differentiation is not taking place twice with respect to an entry of σ, then at least once differentiation has to take place with respect to an entry of τ or d. In this case equation (25) implies that:

$$\partial^2_{i,j}L_T(\alpha) = \mathrm{tr}[(\Sigma^0)^{-1}\frac{2}{T}\sum_{t=1}^{T}(\partial_i\varepsilon_t(\alpha))(\partial_j\varepsilon_t(\alpha))'] + o_P(T^{N_u/2}) \tag{26}$$

The above equation (26) is the starting point for the further considerations.

If differentiation takes place twice with respect to an entry of τ_{st}, then all quantities in the above equation are stationary, see also the previous lemma. In this case convergence to a constant matrix follows.

If differentiation takes place once with respect to an entry of τ_{st} and once with respect to an entry of d, convergence of $\partial^2_{i,j}L_T(\bar{\alpha}) \to 0$ follows.

If differentiation takes place twice with respect to an entry in d, the relevant term is given by $\mathrm{tr}[(\Sigma^0)^{-1}\frac{2}{T}\sum_{t=1}^{T}(k^{-1}(z,\bar{\tau})\partial_i Ds_t)(k^{-1}(z,\bar{\tau})\partial_j Ds_t)']$ This directly implies that the asymptotic entry in the (limit of the) Hessian is only non-zero, if both entries of d with respect to which differentiation takes place correspond to elements in the same column of D, q say. For $z_q \neq \pm 1$ the corresponding limit block in the Hessian is in this case given by

$$\begin{bmatrix} 4(k^0(\overline{z_q},\tau^0)^{-1})'(\Sigma^0)^{-1}k^0(\overline{z_q},\tau^0)^{-1} & 0 \\ 0 & 4(k^0(\overline{z_q},\tau^0)^{-1})'(\Sigma^0)^{-1}k^0(\overline{z_q},\tau^0)^{-1} \end{bmatrix}$$

For $z_q = \pm 1$ the block is equal to $2(k^0(\overline{z_q},\tau^0)^{-1})'(\Sigma^0)^{-1}k^0(\overline{z_q},\tau^0)^{-1}$. Next consider differentiation once with respect to an entry in τ_u and once with respect to an entry of τ_{st}. The corresponding entry is then the sum of a product of a stationary process with an integrated process (integrated of order 1 at the corresponding unit root). The normalization factor $T^{-3/2}$, with which these elements are

scaled, then implies convergence to zero of the scaled quantity.
Only the cases twice differentiating with respect to an entry in τ_u, and differentiating once with respect to an entry of τ_u and once with respect to an entry of d are left to be examined. Note in this respect first that if in differentiating twice with respect to entries of τ_u, the two parameters correspond to different unit roots, it holds that $T^{-1}\partial_{i,j}^2 L_T(\bar{\alpha}_T) \to 0$. This follows from observing that $\partial_i \varepsilon_t(\alpha) = dk_i(z,\tau)x_t(z_k)$, where we assume that the entry with respect to which differentiation takes place corresponds to the pair of complex conjugate unit roots $z_k, \overline{z}_k$ or to $z_k = \pm 1$ and $x_t(z_k)$ denotes as before in Lemma 8 the state vector in the real valued canonical form. The above expression for the partial derivative can directly be investigated using Lemma 3, item iv). The lemma provides similarly the result for the case that both i and j correspond to the same unit root or pair of unit roots:

$$T^{-1}\partial_{i,j}^2 L_T(\bar{\alpha}_T) \xrightarrow{d} \operatorname{tr}\left[2\mathcal{R}\left\{dk_i(\overline{z_k},\tau)Z(z_k)dk_j(\overline{z_k},\tau)'\right\}\right]/\delta_k^2$$

Finally, consider the last possible combination. First differentiation with respect to an entry of τ_u and then with respect to an entry of d. Here we have to distinguish two cases, whether j corresponds to an element of the j-th column of D_r or to an element in the j-th column of D_i (which is corresponding to $s_{t,j}^1 = z_j^t$). Suppose for the moment that it corresponds to an element of D_r, and let i denote a component of τ_u that corresponds to unit root z_i. Then we obtain

$$T^{-1/2}\partial_{i,j}^2 L_T(\bar{\alpha}_T) \xrightarrow{d} 2\mathcal{R}\left((k^0(\overline{z_i},\tau^0)^{-1})'(\Sigma^0)^{-1}dk_i(\overline{z_i},\tau^0)Y(z_i)(\Sigma^0)^{-1}\right)/\delta_i^2$$

for $z_i = z_j$ and zero else. The expression for an element of d corresponding to an entry in the j-th column of D_i is the same, except for replacing $\mathcal{R}$ by $\mathcal{I}$.
It remains to analyze the nonsingularity properties of Z. When D is not estimated, the block-diagonality of the asymptotic Hessian implies that it is sufficient to treat the blocks Z_u and Z_{st} separately. When D is estimated, it is sufficient to investigate Z_D and $Z_u - Y_D' Z_D^{-1} Y_D$.
Consider the block corresponding to τ_{st} first: This block converges in fact to a constant matrix, i.e. asymptotic nonsingularity is shown, if the limiting matrix is nonsingular. For the part of τ_{st} corresponding to the parameters for $k_{st}(z)$ this follows again from standard theory for stationary processes. Thus, only the derivatives corresponding to the parameters for B_u have to be analyzed:

$$\partial_i \varepsilon_t(\alpha^0) = -C\partial_i x_t(\tau^0), \partial_i x_{t+1}(\tau^0) = (A - BC)\partial_i x_t(\tau^0) + \partial_i B \varepsilon_t$$

The proof is indirect: If the matrix corresponding to B_u were singular, there existed a vector $x = [x_1, \ldots, x_v]'$ such that

$$0 = \sum_{r,s=1}^{v} x_r x_s \operatorname{tr}[\Sigma^{-1}\mathbb{E}\partial_s\varepsilon_t(\alpha^0)\partial_r\varepsilon_t(\alpha^0)'] = \operatorname{tr}[\Sigma^{-1}\mathbb{E}\sum_{r=1}^{v} x_r\partial_r\varepsilon_t(\alpha^0)\sum_{s=1}^{v} x_s\partial_s\varepsilon_t(\alpha^0)]$$

denoting the components of τ corresponding to B_u with $1, \ldots, v$ for some integer v. This implies that $\sum_r x_r \partial_r \varepsilon_t(\alpha^0)$ is equal to zero and thus that the filters for generating the score are linearly dependent. This however cannot be the case, as the parameter vector α^0 corresponds by assumption to a minimal system. Thus, the block corresponding to B_u is asymptotically nonsingular. The same argument also shows asymptotic nonsingularity of the whole block corresponding to τ_{st}, because there are also no linear dependencies between the filters corresponding to derivatives with respect to entries in B_u and filters corresponding to derivatives with respect to entries in the stationary part $k_{st}(z)$.
Finally consider the same kind of argument for the part of the second order derivative of the likelihood function corresponding to τ_u. It has been shown above that the essential term in equation (24) is

$$\begin{aligned} \partial_{i,j}^2 Q_T^u &= \sum_{t=1}^{T} \operatorname{tr}[\Sigma^{-1}\partial_i\varepsilon_t(\alpha^0)\partial_j\varepsilon_t(\alpha^0)'] + o_P(1) \\ \partial_i\varepsilon_t &= dk_i(z,\alpha^0)x_t^u = -k(z,\tau^0)^{-1}\partial_i C_i x_t(z_i) \end{aligned}$$

This matrix is block-diagonal, with the blocks corresponding to the different unit roots. Therefore again the crucial fact to prove, is the linear independence of the filters $dk_i(z,\tau^0)$ for all coordinates in

τ_u corresponding to $z_k = e^{i\omega_k}$ for all k. Remember that $k^{-1}(e^{i\omega_k}, \tau^0)C_k^0 = 0$. Therefore, a necessary condition is that the derivative of C_k does not lie in the space spanned by the columns of C_k, and that the set of partial derivatives with respect to all different parameters is linearly independent. This is ensured by the specific parameterization of C_k as described in Appendix B.
The expressions given for Z_u, Y_D and Z_D in the theorem directly show that Z_u and $Z_u - Y_D' Z_D^\dagger Y_D$ have the same structure, where only $Z(z_k)$ is replaced by

$$\begin{aligned}
Z(z_k) - Y(z_k)Y(z_k)' &= \delta_k^2 B_k \int_0^1 W_k W_k' B_k' - \delta_k^2 B_k \int_0^1 W_k \int_0^1 W_k' B_k' \\
&= \delta_k^2 B_k \left(\int_0^1 W_k W_k{}' - \int_0^1 W_k \int_0^1 W_k{}' \right) B_k' \\
&= \delta_k^2 B_k \int_0^1 \left(W_k - \int_0^1 W_k \right) \int_0^1 \left(W_k - \int_0^1 W_k \right)' B_k'
\end{aligned}$$

It follows immediately that this matrix is positive with probability one. This concludes the proof. □

Combining the results of the previous two lemmata, the asymptotic distributions of $T(\hat{\tau}_u - \tau_u^0)$ and $\sqrt{T}(\hat{\tau}_{st} - \tau_{st}^0)$ follow immediately: In any case $\sqrt{T}(\hat{\tau}_{st} - \tau_{st}^0) \xrightarrow{d} \mathcal{N}(0, Z_{st}^{-1} V_{st} Z_{st}^{-1})$. If no deterministic terms are included in the estimation it follows, that the sub-block of $\hat{\tau}_u$ corresponding to the unit root z_k has the following limiting distribution:

$$\left\{ \mathrm{tr} \left[\partial_j C_k' (k^0(\overline{z_k}, \tau^0)^{-1})' (\Sigma^0)^{-1} k^0(\overline{z_k}, \tau^0)^{-1} \partial_i C_k Z(z_k) \right] \right\}^{-1} \left\{ \mathrm{tr}[(\Sigma^0)^{-1} \left(dk_i(\bar{z}_k, \tau^0) X(\omega_k) \right)] \right\}$$

for $z_k = \pm 1$ and

$$\mathcal{R} \left\{ \mathrm{tr} \left[\partial_j C_k' (k^0(\overline{z_k}, \tau^0)^{-1})' (\Sigma^0)^{-1} k^0(\overline{z_k}, \tau^0)^{-1} \partial_i C_k Z(z_k) \right] \right\}^{-1} \mathcal{R} \left\{ \mathrm{tr}[(\Sigma^0)^{-1} \left(dk_i(\bar{z}_k, \tau^0) X(\omega_k) \right)] \right\}$$

else. Here only the typical elements of the respective matrices have been given. When the term Ds_t is included in the estimation, $Z(z_k)$ has to be replaced by $Z(z_k) - Y(z_k)Y(z_k)'$ and $X(z_k)$ has to be replaced by

$$X(z_k) - Y(z_k)W_k(1)' = \delta_k^2 B_k \int_0^1 W_k dW_k' - \delta_k^2 B_k \int_0^1 W_k \int_0^1 1 dW_k' = \delta_k^2 \int_0^1 B_k (W_k - \int_0^1 W_k) dW_k'$$

as follows from straightforward computations. This finally concludes the proof of Theorem 2.

Appendix B

In the construction of the canonical form the matrices C_k are restricted to be orthonormal and positive lower triangular matrices. In this appendix one possibility to parameterize matrices that fulfill these two restrictions is presented. The p.l.t. structure introduces additional integer constraints, the row indices for the first non-zero element in each column. In the parameterization we take these indices to be as given, hence we present a real valued parameterization of the set of all matrices $C \in \mathbb{C}^{s \times c}$ such that $C'C = I_c$ and C is positive lower triangular with indices $1 \leq i_1 < \cdots < i_c \leq s$. Note again that we only present one possibility. The convergence proof of Theorem 1 is performed parameterization free, hence any other suitable preferred parameterization may be employed instead.

Lemma 10 *For each matrix $C \in \mathbb{C}^{s \times c}, C = [c_1, \ldots, c_c], C'C = I_c$, which is in p.l.t. form with indices $1 \leq i_1 < \cdots < i_c$ a real parameter vector $x \in \mathbb{R}^d$ of dimension $d = \sum_{j=1}^c (2s - 2d(i_j)) - 2c$ is given by the following parameters: Let $t' = [t_1', \ldots, t_c'], t_j \in \mathbb{R}^{2s-2d(i_j)-1}$, then $c_j = Q_j f(t_j)$, where Q_j denotes the unique p.l.t. representation of the unitary complement of $[0^{s-i_j+1, i_j-1}, I_{s-i_j+1}][c_1, \ldots, c_{j-1}]$. Here $d(i_j) = i_j + rank\left([0^{s-i_j, i_j-1}, I_{s-i_j+1}][c_1, \ldots, c_{j-1}]\right)$ and $f(t_j) = f^r(t_j) + i f^i(t_j)$, where $[f^r(t_j)', f^i(t_j)']'$ is a real unit norm vector, parameterized using the real parameter vector t_j using e.g. stereographic projections. Furthermore, if e_1 denotes the first vector of the canonical basis, then $e_1' f^r(t_j) > 0$ and $e_1' f^i(t_j) = 0$ holds.*

The parameterization of a real valued d-dimensional unit vector vector requires $(d-1)$ parameters. The additional restriction on the first entry of $f^i(t_j)$ saves one additional parameter. Taking the first entry in $f^t(t_j) > 0$ implies that this set can be parameterized by using the stereographic projection with only one chart.
A problem with this type of parameterization is the use of the p.l.t. form for the recursively computed complements, which introduces further integer parameters: The first problem is that the rank of the matrix $[0^{s-i_j+1,i_j-1}, I_{s-i_j+1}][c_1, \dots, c_{j-1}]$ need not be maximal, but could be less than $(j-1)$. This is taken account of by including the integer $d(i_j)$. Secondly the specification of the p.l.t. structure requires a further integer valued parameter, which possibly complicates the actual implementation of the procedure, due to the recursive nature of the specification of these integers. On the subsets with a constant set of indices i_j, i.e. with constant p.l.t. structure, the parameterization is continuous and differentiable with differentiable inverse.
A disadvantage of the proposed parameterization is its recursive nature, which complicates the calculation of the gradient vector (with respect to the entries in the parameter vector). Here again it is a nice feature that the consistency proof is parameterization independent, hence some other parameterization with possibly more convenient computational properties may be found and used.

References

Aoki, M. (1990). *State Space Modelling of Time Series*. Springer, New York.

Aoki, M. & A. Havenner, eds. (1997). *Applications of Computer Aided Time Series Modelling*. Lecture Notes in Statistics. Springer, New York.

Bauer, D. & M. Wagner (2002a). Estimating Cointegrated Systems Using Subspace Algorithms. Forthcoming in *Journal of Econometrics*.

Bauer, D. & M. Wagner (2002b). A Canonical Form for Unit Root Analysis in the State Space Framework. Submitted to *Journal of Econometrics*.

Bauer, D. & M. Wagner (2002c). On Polynomial Cointegration in the State Space Framework. Submitted to *Econometric Theory*.

Chan, N.H. & C.Z. Wei (1988). Limiting Distributions of Least-Squares Estimates of Unstable Autoregressive Processes. *Annals of Statistics* **16**, 367–401.

Gregoir, S. (1999a). Multivariate Time Series with Various Hidden Unit Roots, Part I: Integral Operator Algebra and Representation Theorem. *Econometric Theory* **15**, 435–468.

Gregoir, S. (1999b). Multivariate Time Series with Various Hidden Unit Roots, Part II: Estimation and Test. *Econometric Theory* **15**, 469–518.

Hannan, E. & M. Deistler (1988). *The Statistical Theory of Linear Systems*. Wiley, New York.

Hazewinkel, M. & R.E. Kalman (1976). Invariants, Canonical Forms and Moduli for Linear, Constant, Finite Dimensional, Dynamical Systems. In: *Lecture Notes in Economics and Mathematical Systems* (G. Marchesini and S.K. Bitter, eds.), 48–60. Springer, New York.

Johansen, S. & E. Schaumburg (1999). Likelihood Analysis of Seasonal Cointegration. *Journal of Econometrics* **88**, 301–339.

Lee, H.S. (1992). Maximum Likelihood Inference on Cointegration and Seasonal Cointegration. *Journal of Econometrics* **54**, 1–47.

Meyer, C.D. (2000). *Matrix Analysis and Applied Linear Algebra.* SIAM, Philadelphia.

Ober, R. (1996). Balanced Canonical Forms. In: *Identification, Adaptation, Learning* (S. Bittanti and G. Picci, eds.), 120–183. Springer.

Phillips, P.C.B. (1991a). Optimal Inference in Cointegrated Systems. *Econometrica* **59**, 283–306.

Phillips, P.C.B. (1991b). Spectral Regression for Cointegrated Time Series. In: *Nonparametric and Semiparametric Methods in Economics and Statistics* (W. Barnett, Ed.), 413–436. Cambridge University Press.

Saikkonen, P. (1992). Estimation and Testing of Cointegrated Systems by an Autoregressive Approximation. *Econometric Theory* **8**, 1–27.

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.

Saikkonen, P. (1995). Problems with the Asymptotic Theory of Maximum Likelihood Estimation in Integrated and Cointegrated Systems. *Econometric Theory* **11**, 888–911.

Saikkonen, P. & R. Luukkonen (1997). Testing Cointegration in Infinite Order Vector Autoregressive Processes. *Journal of Econometrics* **81**, 93–126.

Yap, S.F. & G.C. Reinsel (1995). Estimating and Testing for Unit Roots in a Partially Nonstationary Vector Autoregressive Moving Average Model. *Journal of the American Statistical Association* **90**, 253–267.