JOURNAL OF TIME SERIES ANALYSIS
J. Time Ser. Anal. **43**: 178–196 (2022)
Published online 20 July 2021 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1111/jtsa.12607

ORIGINAL ARTICLE

ON CAUSAL AND NON-CAUSAL COINTEGRATED VECTOR AUTOREGRESSIVE TIME SERIES

ANDERS RYGH SWENSEN*
Department of Mathematics, University of Oslo, Oslo, Norway

Previous treatments of multivariate non-causal time series have assumed stationarity. In this article, we consider integrated processes in a non-causal setting. We generalize the Johansen–Granger representation for causal vector autoregressive (VAR) models to allow for dependence on future errors and discuss how the parameters can be estimated. The asymptotic distribution of the trace statistic is also considered. Some Monte Carlo simulations are presented.

Received 13 October 2020; Accepted 11 June 2021

Keywords: Vector autoregressive models; non-causal models; cointegration

JEL. C32.

MOS subject classification: 62M10; 62M15; 91B84

1. INTRODUCTION

Vector autoregressive, VAR, models are one of the basic tools for analyzing macroeconomic time series. In such time series trends are often observed, and tools for handling such behavior are necessary. Cointegration has played an important role in this respect since it was introduced in the seminal paper by Engle and Granger (1987). One of the implications is that under some regularity conditions there exists a Johansen–Granger representation of a VAR model which means that it is possible to decompose the time series into a random walk part, a stationary part, a non-stochastic part and a part depending on an initial condition. Certain linear combinations of the components of the levels of the time series are stationary and these linear combinations are the cointegration vectors.

The regularity conditions needed for a Johansen–Granger representation ensure that the stationary part is causal, that is, the observations depend on only past and present errors of the VAR. However, this is not the only way to obtain stationarity. In the so-called non-causal and mixed causal non-causal models the observations also depend on future errors. Earlier contributions to the literature on non-causal models are Breidt *et al.* (1991), Andrews *et al.* (2006) and the monograph by Rosenblatt (2000) which all mainly considered univariate time series models. Recently also multivariate models have attracted attention, see Lanne and Saikkonen (2013), Cubadda *et al.* (2019), Gouriéroux and Jasiak (2017), Davis and Song (2020) and Cavaliere *et al.* (2020). For multivariate nonstationary and non-causal time series the question arises whether there is a formulation such that the important property of cointegration implied by a Johansen–Granger representation is retained and analysis based on levels of the time series is permitted.

The models allowing for non-causality which we will consider can be seen as an extension of the stable causal VAR model. Then the roots of the determinant of the autoregressive polynomial are all located outside the unit circle, $\{z : |z| = 1\}$. Requiring only that the determinant of the autoregressive polynomial is non-zero for values

*Correspondence to: Anders Rygh Swensen, Department of Mathematics, University of Oslo, P.O. Box 1053 Blindern, 0316 Oslo, Norway.
E-mail: swensen@math.uio.no

at the unit circle is sufficient to ensure existence of stationary processes, causal or non-causal. If all the roots are inside the unit circle, the process is purely non-causal and if there are roots both inside and outside the unit circle mixed causal or non-causal process is used as denomination.

The well-known integrated and cointegrated processes can be seen as another extension of the stable causal VAR models by permitting also one or more root of the determinant of the characteristic polynomial exactly at 1 in addition to those located outside the closed unit disk.

We shall in this article address the situation where both extensions of the stable VAR models are possible. Then the determinant of the characteristic polynomial of the VAR model can have roots outside and inside the unit circle but also some roots exactly at $z = 1$. Previous studies of reduced rank VAR models allowing roots inside the unit circle have assumed causality in addition, which means that the process is explosive. Johansen (2009) and Nielsen (2010) studied situations where unstable roots are present. Engsted and Nielsen (2012) used such processes to model bubble behavior. Another example where explosive processes arise is in using bootstrap methods to find the appropriate reduced rank r. The usual procedure consists of fitting models of increasing rank. For some estimation methods it may happen that although the data generating model may have a characteristic polynomial where the determinant of the characteristic polynomial is non-zero inside the unit circle, some of the fitted models may not, see, for example, Cavaliere *et al.* (2012) and Swensen (2006). The alternative to introducing explosive processes is to allow for non-causality. This has been done for unit root testing by Saikkonen and Sandberg (2016). We will consider the multivariate situation.

In the present article we show that a Johansen–Granger representation also exists when some roots of the determinant of the characteristic polynomial of the VAR have modulus less than 1 and the only roots with modulus 1 are exactly at $z = 1$. Then the stationary part is no longer causal and may depend on future random shocks.

When autoregressive models of this type are fitted to an observed time series a natural question is how many roots of the determinant of the characteristic polynomial of the VAR are located inside the unit circle. This can be answered once the coefficients of the VAR-model have been estimated. To find out how that can be done is therefore important.

An example where such results can be employed is described in the article by Lanne and Saikkonen (2013) where a non-causal VAR model was introduced to analyze the expectation hypothesis of the term structure of interest rates. In the article a bivariate time series consisting of 6-month and 5-year interest rate was considered using the change in the 6-month rate and the spread between the 6-month and 5-year interest rates. Being able to employ the levels directly and not rely on a transformation to obtain stationarity will be an advantage. For example, one can investigate whether the transformation to stationarity using difference and spread is appropriate.

There are a couple of special features which are prominent for the full rank stationary case which also have implications when the rank is reduced and which are worth mentioning. One is how the dependence of future shocks shall be interpreted. Non-causal models take the uncertainty of future errors, not only present and past errors, into account. This can be a clear advantage in some situations. Gouriéroux and Zakoian (2017), section 3.3, stress that such models with error distributions having fat tails can describe the local explosive behavior often found in economic and financial time series. Taking into account non-causality can be useful also when fitting causal models. Lanne and Saikkonen (2013) pointed out that this is the case when only a subset of the variables which are causally generated is modeled. More specifically, building on a result of Johansen and Juselius (2014), they explained that a linear combination, Y_t, of such a subset may have a one sided representation, $Y_t = \Sigma_{i=0}^{\infty} \Xi_i \epsilon_{t-i}^{y}$ where the errors ϵ_t^y are correlated with $Y_{t-i},\ i = 1, \ldots$. Such dependence is typical for non-causal time series having a two-sided representation. Hence, if non-causal models provide better fit than causal models, omissions of this kind may be the explanation.

The other aspect is identifiability. To a VAR model where some of the roots of the determinant of the characteristic polynomial are inside the unit circle, it is possible to specify a new VAR with a different variance of the errors, $Var(\epsilon_t)$, which have all the roots outside the unit circle, that is, a causal model. It can be shown that the two models have the same first and secondary moment structure. The distribution of a Gaussian model is defined by this structure. Two Gaussian models with the same first and secondary moment structure will therefore have the same distribution. To ensure identifiability in Gaussian models the parameter space is restricted such that only causal models are permitted. For non-causal model to be identifiable only non-Gaussian models are permitted and some additional restrictions must be imposed.

The article is organized as follows. In the next section we prove a Johansen–Granger representation for a non-causal VAR model. For the case where there are no deterministic terms we discuss in Sections 3 and 4 how the unknown parameters can be estimated and find the asymptotic distribution of the trace test for determining the rank. Section 5 contains results from some Monte Carlo simulations.

Some additional results can be found in the online supporting information.

2. A JOHANSEN–GRANGER REPRESENTATION THEOREM

The vector autoregressive, VAR, model of dimension p and order k is defined by the recursion

$$X_t = A_1X_{t-1} + \cdots + A_kX_{t-k} + \epsilon_t, \tag{1}$$

where $\{\epsilon_t\}$ is a series of uncorrelated random variables with expectation zero and finite second-order moment. The traditional stability requirement, see Hannan and Deistler (1988) and Lütkepohl (2005), that the determinant of the autoregressive or characteristic polynomial, $A(z) = I - A_1z - \cdots - A_kz^k$, is non-zero on the closed unit disk, $\{z : |z| \leq 1\}$, is sufficient to ensure that $\{X_t\}_{t=-\infty}^{\infty}$ is stationary and can be expressed as a linear filter of the present and past values of the variables ϵ_t, that is,

$$X_t = \Sigma_{j=0}^{\infty} C_j\epsilon_{t-j},\ t = 0, \pm1, \pm2, \ldots \text{ where } \Sigma_{j=0}^{\infty} tr(C_j'C_j) < \infty. \tag{2}$$

VAR models satisfying the stability requirement have been extensively studied and are an important ingredient in applications in many fields such as empirical macroeconomics, engineering and climate research, just to mention a few. Important references in addition to the mentioned Hannan and Deistler (1988) and Lütkepohl (2005) are Hannan (1970), Brockwell and Davis (1991) and Hendry (1995).

For models allowing roots of the determinant of the characteristic polynomial on the unit circle the alternative vector equilibrium, VECM, formulation is useful. The recursion (1) can be reparameterized as

$$\Delta X_t = \Pi X_{t-1} + \Gamma_1\Delta X_{t-1} + \cdots + \Gamma_{k-1}\Delta X_{t-k+1} + \Phi D_t + \epsilon_t, \tag{3}$$

where a deterministic term D_t has been added. When the rank r of the matrix Π is reduced, $0 \leq r < p$, $\det A(1) = 0$ and it follows from the Johansen–Granger theorem, see Johansen (1995), that under some regularity conditions, $X_t,\ t = k+1, k+2, \ldots$ can be decomposed into a random walk, a stationary part, a non-stochastic part and a term depending on the initial values $X_1, \ldots, X_k$.

A Johansen–Granger representation allowing for non-causality can be proved under assumptions which are quite similar to those used in the causal situation. In fact, the only change is that solutions of $\det A(z) = 0$ which have moduli strictly less than 1 are permitted.

Assumption 1. The recursion (3) satisfies the following conditions

(i) the determinant of the characteristic polynomial has roots which are exactly at 1 or have moduli which are either strictly less than 1 or strictly larger than 1, that is $\det A(z) = 0$ implies $z = 1$ or $|z| \neq 1$,
(ii) the matrix $\Pi = \alpha\beta'$ where matrices α and β have full rank r with $0 \leq r < p$,
(iii) the matrix

$$\alpha_\perp'\Gamma\beta_\perp$$

has full rank $p - r$, where $\Gamma = I_p - \Gamma_1 - \cdots - \Gamma_{k-1}$.

   DOI: 10.1111/jtsa.12607

The following notation is used. If an $m \times n$ matrix a, where $n \leq m$, has full rank, $a_{\perp}$ denotes an $m \times (m-n)$ matrix of full rank such that $a'_{\perp}a = 0$. The matrix $a(a'a)^{-1}$ is defined as $\bar{a}$, so that $a'\bar{a} = I_n$ and $\bar{a}a'$ is the projection matrix on the space spanned by the columns of a.

The following assumptions on the distribution of the errors and the behavior of the deterministic terms are also needed.

Assumption 2. The errors ϵ_t are i.i.d. random variables with expectation 0 and covariance matrix Ω.

Assumption 3. There exist constants a and b such that the deterministic term, D_t, satisfies $|D_t| < a + |t|^b$.

The idea is to express, following Hansen (2005), the model (3) in a companion form, for suitable $pk \times l$, $l = p(k-1) + r$ matrices α^* and β^*, as

$$\Delta X_t^* = \alpha^* \beta^{*\prime} X_{t-1}^* + \Phi_t^* + \epsilon_t^*, \; t = k+1, \dots \tag{4}$$

where $X_t^* = (X'_t, \dots, X'_{t-k+1})'$, $\epsilon_t^* = (\epsilon'_t, 0, \dots, 0)'$ and $\Phi_t^* = ((\Phi D_t)', 0, \dots, 0)'$. Multiplying both sides of (4) with $\beta^{*\prime}$ and rearranging yields

$$\beta^{*\prime} X_t^* = (I + \beta^{*\prime}\alpha^*)\beta^{*\prime} X_{t-1}^* + \beta^{*\prime}(\epsilon_t^* + \Phi_t^*). \tag{5}$$

Under Assumption 1, as shown in Appendix A, there exist nonsingular matrices M, G_1, G_2 such that

$$I + \beta^{*\prime}\alpha^* = MGM^{-1} = M \begin{pmatrix} G_1 & 0 & 0 \\ 0 & G_2 & 0 \\ 0 & 0 & 0 \end{pmatrix} M^{-1}$$

where all the eigenvalues of G_1 have modulus less than 1, all the eigenvalues of G_2 have modulus larger than 1 and the lower right block is present only when the matrix $I + \beta^{*\prime}\alpha^*$ is singular.

In Appendix A and the supporting information one can find a proof of the following.

Proposition 1. Under Assumptions 1–3 and with the matrices M, G_1 and G_2 as described above, X_t can be represented as

$$X_t = C \sum_{s=k+1}^{t} (\epsilon_s + \Phi D_s) + \sum_{s=-\infty}^{\infty} C_s(\epsilon_{t-s} + \Phi D_{t-s}) + A, t = k+1, \dots \tag{6}$$

where $C = \beta_{\perp}(\alpha'_{\perp}\Gamma\beta_{\perp})^{-1}\alpha'_{\perp}$ and $C_s = FMC_s^* M^{-1}B$ with $B = (\beta, I, 0, \dots, 0)'$ if $k > 1$. If $k = 1$, $B = \beta'$. With $\Gamma_i^* = \Gamma_i + \cdots + \Gamma_{k-1}$ the matrices F and C_s^* are $F = ((I - C\Gamma)\bar{\beta}, -C\Gamma_1^*, \dots, -C\Gamma_{k-1}^*)$ and

$$C_s^* = \begin{pmatrix} G_1^s & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \; when \;\; s \geq 0 \; \text{ and } C_s^* = -\begin{pmatrix} 0 & 0 & 0 \\ 0 & G_2^s & 0 \\ 0 & 0 & 0 \end{pmatrix} \; when \;\; s < 0.$$

The term $A = C(X_k - \Gamma_1 X_{k-1} - \cdots - \Gamma_{k-1}X_1)$ depends only on the initial values and $\beta' A = 0$.

Conversely, if $I + \beta^{*\prime}\alpha^*$ is non-singular, a process satisfying (6) where $C_s = FMC_s^* M^{-1}B$, must satisfy the recursion defined in (3) for t = k+1,….

To illustrate the implications of Proposition 1 we consider the following simple example where there is a double root of $\det A(z) = 0$ at 1 and one outside or inside the unit circle.

Example 1. Let $p = 3$, $k = 1$ and $r = 1$ and consider the three dimensional VAR model

$$\Delta X_t = \alpha\beta' X_{t-1} + \epsilon_t,\ t = 2, 3, \ldots \tag{7}$$

Then

$$\beta^* = \beta,\ \alpha^* = \alpha \text{ and } I_1 + \beta^{*\prime}\alpha^* = 1 + \beta'\alpha.$$

Also $\beta' X_t$ satisfies $\beta' X_t = (1 + \beta'\alpha)\beta' X_{t-1} + \beta'\epsilon_t$. The determinant of the characteristic polynomial is a third order polynomial. We consider a simple situation where $\beta = (1, 0, 1)'$ and $\alpha = (a, a, a)'$. Then

$$I_1 + \beta^{*\prime}\alpha^* = 1 + 2a.$$

The determinant of the characteristic polynomial of (7) is $\det A(z) = (1 - z)^2(1 - z - 2az)$ which has a double root at 1 and a single root at $1/(1 + 2a)$. With

$$\beta_\perp = \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ 0 & -1 \end{pmatrix} \text{ and } \alpha_\perp = \begin{pmatrix} a & a \\ 0 & -2a \\ -a & a \end{pmatrix}$$

the matrix C is

$$C = \beta_\perp(\alpha'_\perp\Gamma\beta_\perp)^{-1}\alpha'_\perp = \begin{pmatrix} 1/2 & 0 & -1/2 \\ -1/2 & 1 & -1/2 \\ -1/2 & 0 & 1/2 \end{pmatrix}$$

since in this case $\Gamma = I_3$. Then the non-zero coefficients are $C_s = FMC^*_s M^{-1}\beta' = (I - C)\bar{\beta}\beta'(1 + 2a)^s$ when $s \geq 0$ and $|1 + 2a| < 1$ and $C_s = -(I - C)\bar{\beta}\beta' 1(1 + 2a)^s$ when $s < 0$ and $|1 + 2a| > 1$. Since

$$(I - C)\bar{\beta}\beta' = \begin{pmatrix} 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 \end{pmatrix}$$

X_t can be expressed as

$$X_t = \frac{1}{2}\sum_{s=2}^{t} \begin{pmatrix} \epsilon_{1,s} - \epsilon_{3,s} \\ -\epsilon_{1,s} + 2\epsilon_{2,s} - \epsilon_{3,s} \\ -\epsilon_{1,s} + \epsilon_{3,s} \end{pmatrix} + \begin{pmatrix} 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 \end{pmatrix} S_t + \begin{pmatrix} X_{1,1} - X_{3,1} \\ -X_{1,1} + 2X_{2,1} - X_{3,1} \\ -X_{1,1} + X_{3,1} \end{pmatrix} /2$$

where

$$S_t = \sum_{s=0}^{\infty}(1 + 2a)^s \begin{pmatrix} \epsilon_{1,t-s} \\ \epsilon_{2,t-s} \\ \epsilon_{3,t-s} \end{pmatrix} \text{ when } |1 + 2a| < 1$$

and

$$S_t = -\sum_{s=1}^{\infty}(1 + 2a)^{-s} \begin{pmatrix} \epsilon_{1,t+s} \\ \epsilon_{2,t+s} \\ \epsilon_{3,t+s} \end{pmatrix} \text{ when } |1 + 2a| > 1.$$

X_t is an I(1) process since ΔX_t is stationary as one can see by direct subtraction. Also $\beta' X_t = (1, 0, 1)X_t = X_{1,t} + X_{3,t}$ is stationary since the random walk part disappears and only the linear filter where the coefficients are decaying exponentially remains.

As one can see, according to whether the root $1/(1 + 2a)$ of the determinant of the characteristic polynomial is outside or inside the unit circle the stationary part is causal or purely non-causal. For more complex situations where there are roots both outside and inside the unit circle the stationary part will be a combination of causal and non-causal processes.

If we let $L^{-1}x_t = x_{t+1}$ in a sequence $\{x_t\}_{-\infty}^{\infty}$ the following Corollary is immediate. One can find a definition of Laurent series in Brockwell and Davis (1991, p. 88).

Corollary 1. Under Assumptions 1–3 X_t can be represented as

$$X_t = C \sum_{s=k+1}^{t} (\epsilon_s + \Phi D_s) + C(L)(\epsilon_t + \Phi D_t) + A, t = k + 1, \dots \tag{8}$$

where $C = \beta_\perp(\alpha_\perp' \Gamma \beta_\perp)^{-1}\alpha_\perp'$ and $C(z)$ is a Laurent series converging in an annulus $1 - \delta < |z| < 1 + \delta$ for some $\delta > 0$.

Corollary 2. Consider the recursion (3) for $t = k + 1, \dots, T$. Under Assumptions 1–3 the variables $\tilde{X}_{1,t} = \sum_{s=0}^{\infty} C_s^* M^{-1} B \epsilon_{t-s}$ and $\tilde{X}_{2,t} = \sum_{s=-\infty}^{-1} C_s^* M^{-1} B \epsilon_{t-s}$ satisfy

$$\tilde{X}_{1,t} = \begin{pmatrix} G_1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tilde{X}_{1,t-1} + \begin{pmatrix} I & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} M^{-1} B \epsilon_t, \ t = k + 1, \dots, T$$

$$\tilde{X}_{2,t} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & G_2^{-1} & 0 \\ 0 & 0 & 0 \end{pmatrix} \tilde{X}_{2,t+1} - \begin{pmatrix} 0 & 0 & 0 \\ 0 & G_2^{-1} & 0 \\ 0 & 0 & 0 \end{pmatrix} M^{-1} B \epsilon_{t+1}, \ t = T, \dots, k + 1.$$

Remark 1. Because $\Pi = -A(1) = \alpha\beta'$, the equation $\det A(z) = 0$ has at least $p - r$ roots at $z = 1$, but in combination with the third requirement in Assumption 1 it follows by the argument in Corollary 4.3 in Johansen (1995) that exactly $p - r$ roots are equal to 1.

Remark 2. By fixing T, see Corollary 2, a variety of solutions can be obtained using the method used to prove Proposition 1. In the following we only discuss the solution of the form (6) since considering the others will depend on specifying terminal conditions for $X_{T+1}, \dots, X_{T+k}$.

Remark 3. The variables $\tilde{X}_{1,t}$ and $\tilde{X}_{2,t}$ in Corollary 2 are related to what Gouriéroux and Jasiak (2017, p. 119), call the causal and non-causal components. One can see that an error ϵ_t at time t has a direct effect on the random walk component and the causal part, but lagged one period on the non-causal part.

Remark 4. The time series $\beta' X_t$ is a VAR of order k which can be non-causal. It is well known that such time series are not identified if the process is Gaussian. To ensure identifiability we therefore assume the following.

Assumption 4. The distribution of the errors $\{\epsilon_t\}$ is non-Gaussian with density f_ϵ and the time series $\{\beta' X_t\}$ is identified up to a change in scale and shift in the time origin of the error series.

That two series are not identified aside from change in time of origin and scale essentially means that in two formulations as in (2) with i.i.d. errors, for all t $\tilde{\epsilon}_{t-q} = H\epsilon_t$ and $C_{t-q} = \tilde{C}_t H$ for an integer q and a nonsingular matrix H. Sufficient conditions that ensure that Assumption 4 is fulfilled can be found in Lanne and Saikkonen (2013) and Davis and Song (2020).

3. ESTIMATION OF THE PARAMETERS

Let $X_1, \dots, X_T$ be T observations from the autoregressive model (3) satisfying Assumptions 1, 2 and 4. In the following we only consider the situation where $\Phi = 0$ so Assumption 3 is automatically satisfied. From Proposition 1 it follows that the time series X_t consists of a sum of a random walk component and a stationary process in addition to a term depending on the initial condition. The additional Assumption 4 ensures that X_t is identified when the stationary part is causal or non-causal.

For parameterization of a stationary non-causal VAR model Lanne and Saikkonen (2013) used the formulation

$$A(z) = \tilde{\Pi}(z)\tilde{\Phi}(z^{-1}) \tag{9}$$

where $\tilde{\Pi}$ and $\tilde{\Phi}$ are matrix polynomials. Both $\det \tilde{\Pi}(z) = 0$ and $\det \tilde{\Phi}(z) = 0$ were required to have solutions strictly larger than 1 in absolute value. Davis and Song (2020) studied stationary non-causal time series with the conventional formulation (1) but allowing for solutions of $\det A(z) = 0$ which have moduli strictly less than 1 in addition to moduli strictly larger than one. We will follow their approach, which has the advantage of not needing to specify the orders of $\tilde{\Pi}$ and $\tilde{\Phi}$.

3.1. An Approximate Likelihood

Assume first that β is fixed and normalized by requiring $c'\beta = I$ where c is a known $p \times r$ matrix. Then the process $\beta' X_t$ is stationary and the known results for that situation can be used. Thus, following Davis and Song (2020) we express the distribution of the variables $(X_k^{*\prime}\beta^*, \Delta X_{k+1}^{*\prime}, \dots, \Delta X_T^{*\prime})'$ as a transformation of the errors $\epsilon_t, t = 0, \pm 1, \pm 2 \dots$. It turns out that the influence of the errors before $t = k$ and after $t = T + 1$ can be ignored asymptotically. Some additional notation is necessary for describing the transformation. From the Jordan canonical form theorem there exists a nonsingular matrix M_1 such that $I + \beta^{*\prime}\alpha^* = M_1 J M_1^{-1}$ where $J = J_1 \oplus J_2$, that is, J is block diagonal with diagonal blocks J_1 and J_2 consisting of the canonical blocks with eigenvalues strictly smaller and strictly larger than 1 in modulus respectively. The notation $J(\phi)$, $J_1(\phi)$ and $J_2(\phi)$ will be used when it is useful to emphasize that J, J_1 and J_2 are functions of the autoregressive parameters. Since $M_1^{-1}\beta^{*\prime}X_t^* = JM_1^{-1}\beta^{*\prime}X_{t-1}^* + M_1^{-1}\beta^{*\prime}\epsilon_t^*$, by using the block diagonal structure of J, $V_t^1 = J_1 V_{t-1}^1 + \varepsilon_t^1$ and $V_t^2 = J_2 V_{t-1}^2 + \varepsilon_t^2$ where $M_1^{-1}\beta^{*\prime}X_t^* = V_t = (V_t^{1\prime}, V_t^{2\prime})'$ and $\varepsilon_t = (\varepsilon_t^{1\prime}, \varepsilon_t^{2\prime})' = M_1^{-1}\beta^{*\prime}\epsilon_t^*$. The variables V_t^1 and V_t^2 can be expressed as linear filters with coefficients that are exponentially decaying. They are therefore bounded in probability. The following result is proved in Appendix C.

Proposition 2. Assume that Assumptions 1, 2 and 4 are satisfied. Then there exists a non-singular matrix K with $\det K = \det M_1(-\det J_2^{-1})^{T-k}$ such that

$$\begin{pmatrix} \beta^{*\prime}X_k^* \\ \Delta X_{k+1}^* \\ \vdots \\ \Delta X_T^* \end{pmatrix} = K \begin{pmatrix} V_k^1 \\ \epsilon_{k+1}^* \\ \vdots \\ \epsilon_T^* \\ V_T^2 \end{pmatrix} \tag{10}$$

where $V_t = (V_t^{1\prime}, V_t^{2\prime})' = M_1^{-1}\beta^{*\prime}X_t^*$. V_k^1 depends on $\epsilon_t,\ t = -\infty, \dots, k$ and V_T^2 depends on $\epsilon_t,\ t = T+1, \dots$.

Let $\mathbf{X}$ and $\mathbf{E}$ denote the vectors on the left and right side of (10) respectively and g and h be their densities. The density of $\mathbf{X}$ is therefore $h(K^{-1}\mathbf{x})|\det K^{-1}|$. The components of $\mathbf{E} = K^{-1}\mathbf{X}$ are independent. Denote the parameters by $\boldsymbol{\theta}$ and assume that ϵ_t has density f_ϵ. One can then, as in Davis and Song (2020), consider the point wise approximate log likelihood

$$l_T(\boldsymbol{\theta}) = \sum_{t=k+1}^{T} [\log f_\epsilon(\epsilon_t(\boldsymbol{\phi}); \boldsymbol{\theta}) + \log|\det J_2(\boldsymbol{\phi})|] \tag{11}$$

where $\epsilon_t(\phi) = \Delta X_t - \alpha\beta' X_{t-1} - \Gamma_1 \Delta X_{t-1} - \cdots - \Gamma_{k-1}\Delta X_{t-k+1}$ and the first k observations $X_1, \ldots, X_k$ are considered as given. The approximation consists of ignoring the term $\log\det M_1$ and the contribution from V_k^1 and V_T^2. Asymptotically these terms will not contribute to a log likelihood with $T - k$ terms since they are bounded in probability.

For fixed β the parameters of a general parameter space where the approximate likelihood is defined can be described by dividing $\boldsymbol{\theta}$ into three parts; first, the coefficients in the autoregressive recursion, $\alpha, \Gamma_1, \ldots, \Gamma_{k-1}$ denoted by ϕ; second, the parameters describing the correlation of the errors denoted by σ, that is, the set of symmetric positive definite $p \times p$ matrices; third, the other parameters in the error distribution. They are denoted by ν and are assumed to belong to an open d-dimensional set. The parameters ϕ are a subset of $\mathbb{R}^{pr+(k-1)p^2}$ satisfying the requirements specified in Assumption 1. These requirements define a union of open subsets so ϕ and $\boldsymbol{\theta}$ belong to open sets.

The term $\det J_2(\phi)$ in (11) is the product of the eigenvalues of $I + \beta^{*\prime}\alpha^*$ having moduli larger than 1, so for fixed values of the parameters it can be determined by a procedure computing the eigenvalues. Once the autoregressive parameters are known the eigenvalues defining $\det J_2(\phi)$ are the inverse of the solutions of $\det A(z) = 0$ with moduli less than 1. The number of such solutions determines also the dimension of $J_2(\phi)$. The extra term $(T - k)\log|\det J_2(\phi)|$ in the likelihood when the process is non-causal is positive and increases with the number of solutions of $\det A(z) = 0$ having moduli less than one and also with their distance from the unit circle. However, since $\det A(0) = 1$ the eigenvalues in $\det J_2(\phi)$ must be finite since their inverse values are roots of $\det A(z)$ as pointed out by Hansen (2005) in the proof of his Lemma A.2.

The behavior of $J_2(\phi)$ can be quite complicated when some of the eigenvalues are not simple. However in the case where they are distinct the matrices $J_1 = J_1(\phi)$ and $J_2(\phi)$ are diagonal with the diagonal elements equal to the eigenvalues. Furthermore, the eigenvalues are continuous and differentiable as functions of the matrix entries, see for example, Theorem 6.3.12 in Horn and Johnson (2013). Also the decomposition $I + \beta^{*\prime}\alpha^* = M_1 J M_1^{-1}$ simplifies since $J = J(\phi)$ will be a diagonal matrix with the eigenvalues of $I + \beta^{*\prime}\alpha^*$ as diagonal entries and the eigenvectors as columns of M_1

In Lanne and Saikkonen (2013) a similar approach to derive the distribution of the observations and an approximate likelihood as in (10) and (11) can be found. However due to their use of another parameterization of the process the expression for the approximative likelihood is different.

3.2. A Two-Step Procedure

We first discuss how the maximum likelihood estimators of the elements of $\boldsymbol{\theta}$ can be determined for fixed values of β for a restricted parameter space which is confined to the situation where the eigenvalues of $I + \beta^{*\prime}\alpha^*$, that is, the roots of the determinant of the characteristic polynomial of the autoregressive process $\beta^{*\prime}X_t^*$ of (5), are distinct. Multiplicity of the roots larger than 1 imposes constraints on the autoregressive coefficients. Hence the complement, which defines the region where the eigenvalues are distinct, that is, the restricted parameter space, is a union of open sets.

The distribution of $\{\epsilon_t\}$ is assumed to satisfy the conditions A.1–A.7 in Davis and Song (2020) or the corresponding conditions in Lanne and Saikkonen (2013). In particular we focus on the situation where the distribution is elliptic, that is, $f_\epsilon(x; \boldsymbol{\theta}) = (\det\Sigma)^{-1/2} f(x'\Sigma^{-1}x; \nu)$ where ν is a d-dimensional parameter.

Let now $\boldsymbol{\theta}$ be an interior point of the restricted parameter set. It follows using the arguments in Davis and Song (2020) that the likelihood is differentiable.[1] Therefore there exists a consistent sequence of roots of the approximate maximum likelihood equation, $\frac{\partial}{\partial\theta} l_T(\boldsymbol{\theta}) = 0$, as $T \to \infty$. It can also be proved that these estimators are asymptotically normally distributed, that is, $\sqrt{T}(\hat{\theta}(\beta) - \theta) \xrightarrow{w} N_p(0, I^{-1}(\beta))$ or $\sqrt{T}I^{1/2}(\beta)(\hat{\theta}(\beta) - \theta) \xrightarrow{w} N_p(0, I)$. For verifying that the location of the global maximum converges an alternative argument is necessary, see for example, van der Wart (1998, p. 68). Notice that the maximization to determine the parameters ϕ can be done over the whole space $\mathbb{R}^{pr+(k-1)p^2}$, since under the conditions we have imposed the probability of choosing values in the exceptional subsets will tend to zero. Also these exceptional parts will be very small parts of the set where the likelihood is defined.

Next, consider the case where β is not known. We will consider the situation where a consistent estimator $\hat{\beta}$ is plugged in for β and investigate the asymptotic distribution of this estimator $\hat{\theta}(\hat{\beta})$. By writing

$$\begin{aligned}\sqrt{T}(\hat{\theta}(\hat{\beta}) - \theta) &= \sqrt{T}(\hat{\theta}(\beta) - \theta) + \sqrt{T}(\hat{\theta}(\hat{\beta}) - \hat{\theta}(\beta)),\\ I^{1/2}(\hat{\beta}) &= I^{1/2}(\beta) + (I^{1/2}(\hat{\beta}) - I^{1/2}(\beta))\end{aligned}$$

and multiplying the two expressions we see that a sufficient condition for $\sqrt{T}I^{1/2}(\hat{\beta})(\hat{\theta}(\hat{\beta}) - \theta) \xrightarrow{w} N_p(0, I)$ is that

$$\sqrt{T}(\hat{\theta}(\hat{\beta}) - \hat{\theta}(\beta)) = o_P(1) \text{ and } I^{1/2}(\hat{\beta}) - I^{1/2}(\beta) = o_P(1) \tag{12}$$

and that $\sqrt{T}(\hat{\beta} - \beta) = o_P(1)$ is crucial for (12). Although likely to be valid in many cases a formal proof of (12) is needed for a verification of the plug in procedure.

A candidate for estimating β satisfying $\sqrt{T}(\hat{\beta} - \beta) = o_P(1)$ is based on the common method of solving a generalized eigenvalue problem, which amounts to using the maximum likelihood estimator from the situation where the stationery part is causal and the errors are Gaussian. In the notation of Johansen (1995) model (3) can be written $Z_{0t} = \alpha\beta' Z_{1t} + \Psi Z_{2t} + \epsilon_t$ where $Z_{0t} = \Delta X_t$, $Z_{1t} = X_{t-1}$ and $Z_{2t} = (\Delta X'_{t-1}, \ldots, \Delta X'_{t-k+1})'$. Then for $i, j = 0, 1$ define

$$S_{ij} = \frac{1}{T-k} \sum_{t=k+1}^{T} \left[Z_{it}Z'_{jt} - \left(\sum_{t=k+1}^{T} Z_{it}Z'_{2t}\right)\left(\sum_{t=k+1}^{T} Z_{2t}Z'_{2t}\right)^{-1}\left(\sum_{t=k+1}^{T} Z_{2t}Z'_{jt}\right)\right]. \tag{13}$$

Proposition 3. Let

$$S(\lambda) = \lambda S_{11} - S_{10}S_{00}^{-1}S_{01} \tag{14}$$

and define $\hat{\beta}$ as the r eigenvectors of corresponding to the r largest solutions of $\det S(\lambda) = 0$ normalized as $\hat{\beta}' S_{11}\hat{\beta} = I$. Assume that Assumptions 1 and 2 are satisfied. Then the estimator $\hat{\beta}$ normalized by $c'\hat{\beta} = I$ is consistent and $\sqrt{T}(\hat{\beta} - \beta) = o_P(1)$.

Proof. By inspecting the proof of Lemma 13.1 in Johansen (1995) one can see that the essential element of the proof of the consistency of $\hat{\beta}$ in the causal case is the decomposition of X_t into a random walk and a stationary part. From Proposition 1 the argument remains valid also in the non-causal situation, so β can be estimated as the r eigenvectors of

$$\det S(\lambda) = \det(\lambda S_{11} - S_{10}S_{00}^{-1}S_{01}) = 0 \tag{15}$$

[1] The argument in Davis and Song (2020) relies on differentiating the term $\log\det J_2(\phi)$. A sufficient condition for this to be valid is that the eigenvalues are distinct.

corresponding to the r largest eigenvalues and normalized by $\hat{\beta}' S_{11} \hat{\beta} = I$. It follows from Lemma 6 (i) and (iii) in the supplementary material that $S_{11}\beta = O_P(1)$ and from Lemma 7 (ii) that $S_{10} = O_P(1)$. By pre- and post-multiplying $S(\lambda)$ with $(\beta', \bar{\beta}_\perp' / \sqrt{T})'$ and applying Lemma 7 (i) from the supporting information the solutions of the equation $\det S(\lambda) = 0$ converge to the solutions of

$$\det(\lambda \Sigma_{\beta\beta} - \Sigma_{\beta 0} \Sigma_{00}^{-1} \Sigma_{0\beta}) \det \left(\lambda \bar{\beta}_\perp' C \int_0^1 W_u W_u' \mathrm{d}u C' \bar{\beta}_\perp \right) = 0 \tag{16}$$

where $W_u, 0 \leq u \leq 1$ is a p-dimensional Brownian motion with $Cov(W_u) = u\Omega$ and Σ_{00}, $\Sigma_{\beta 0}' = \Sigma_{0\beta}$ and $\Sigma_{\beta\beta}$ are the limits in probability of S_{00}, $S_{0\beta} = S_{01}\beta$ and $S_{\beta\beta} = \beta' S_{11} \beta$ respectively. Now consider the space spanned by the the r eigenvectors of (15) corresponding to the r largest solutions of $\det S(\lambda) = 0$. Arguing as in Johansen (1995) it follows that this space converges to the space spanned by the r first unit vectors and that $\tilde{\beta} = \hat{\beta}(\bar{\beta}'\hat{\beta})^{-1}$ is a consistent estimator of β and $\sqrt{T}(\tilde{\beta} - \beta) = o_P(1)$. However according to his formula (13.3) $\sqrt{T}(\hat{\beta} - \beta) = (I - \beta' c)\sqrt{T}(\tilde{\beta} - \beta) + O_P(|\tilde{\beta} - \beta|^2) = o_P(1) + O_P(T^{-1/2)} = o_P(1)$. □

A more thorough treatment of other possible estimators for β and their asymptotic distributions is outside the scope of the present article, but is an interesting question.

4. THE ASYMPTOTIC DISTRIBUTION OF THE TRACE TEST

Next, we address the question of how to determine the rank when $\Phi = 0$. The model where the rank is r is denoted by $H(r)$. The trace test of the hypothesis $H(r)$ versus $H(p)$ is based on the statistic $Q_r = -(T-k) \sum_{j=r+1}^{p} \log(1 - \hat{\lambda}_j)$ where $1 > \hat{\lambda}_1 > \cdots > \hat{\lambda}_p$ are the ordered solutions of $\det S(\lambda) = 0$ where $S(\lambda) = \lambda S_{11} - S_{10} S_{00}^{-1} S_{01}$. The hypothesis is rejected for large values of Q_r. For the causal situation the trace test is the likelihood ratio test for testing $H(r)$ versus $H(p)$ when the errors are Gaussian. The argument for the asymptotic distribution in this case depends on a representation of the same type as in Proposition 1, but with $\Sigma_{i=-\infty}^{-1} C_i z^i = 0$, so that the stationary part does not involve future errors.

Without the causality assumption the asymptotic distribution is more complicated but can be found by elaborating on the result of Theorem 11.1 in Johansen (1995). Consider the decomposition of $\det[(\beta, \bar{\beta}_\perp)' S(\lambda)(\beta, \bar{\beta}_\perp)']$ as

$$\det[\beta' S(\lambda)\beta] \det[\bar{\beta}_\perp' \{S(\lambda) - S(\lambda)\beta[\beta' S(\lambda)\beta]^{-1}\beta' S(\lambda)\}\bar{\beta}_\perp].$$

Then if $\rho = T\lambda$ and $T \to \infty$ for λ fixed, $\beta' S(\lambda)\beta = -\Sigma_{\beta 0}\Sigma_{00}^{-1}\Sigma_{0\beta} + o_P(1)$ so the probability limit of $\det[\beta' S(\lambda)\beta]$ is different from zero for all λ. Following Johansen (1995) $\bar{\beta}_\perp' \{S(\lambda) - S(\lambda)\beta[\beta' S(\lambda)\beta]^{-1}\beta' S(\lambda)\}\bar{\beta}_\perp$ equals

$$\rho T^{-1} \bar{\beta}_\perp' S_{11} \bar{\beta}_\perp - \bar{\beta}_\perp' S_{10} N_0 S_{01} \bar{\beta}_\perp + o_P(1)$$

where

$$N_0 = \Sigma_{00}^{-1} - \Sigma_{00}^{-1}\Sigma_{0\beta}[\Sigma_{\beta 0}\Sigma_{00}^{-1}\Sigma_{0\beta}]^{-1}\Sigma_{\beta 0}\Sigma_{00}^{-1}.$$

In the causal case $\alpha' N_0 = 0$ which simplifies the derivation of the asymptotic distribution. More generally the asymptotic distributions of S_{11} and S_{10} are given in Lemma 7 in the supporting information. Recall that $W_u, 0 \leq u \leq 1$ is a p-dimensional Brownian motion with $Cov(W_u) = u\Omega$. Then

$$S_{10} \xrightarrow{w} N_1 - N_2 N_4^{-1} N_3' + C \int_0^1 W_u \mathrm{d}W_u' + C_{-1}\Omega - N_2 N_4^{-1}(C_{-1}' - C_{-2}', \ldots, C_{-k+1}' - C_{-k}')'\Omega \tag{17}$$

DOI: 10.1111/jtsa.12607
Journal of Time Series Analysis published by John Wiley & Sons Ltd.

weakly, where N_1 and N_2 are random matrices distributed as the asymptotic distribution of $\frac{1}{T-k}\sum_{t=k+1}^{T} Z_{1t}Z_{1t}'\beta\alpha'$ and $\frac{1}{T-k}\sum_{t=k+1}^{T} Z_{1t}Z_{2t}'$ respectively. N_3' and N_4 are the limits in probability of $\frac{1}{T-k}\sum_{t=k+1}^{T} Z_{2t}Z_{1t}'\beta\alpha'$ and $\frac{1}{T-k}\sum_{t=k+1}^{T} Z_{2t}Z_{2t}'$ respectively. Expressions for N_1, N_2 and N_3 can be found in Lemma 6 in the supporting information.

Proposition 4. Assume $\Phi = 0$ and that Assumptions 1 and 2 are satisfied. Denoting $\bar{\beta}_{\perp}' C \int_0^1 W_u W_u' \mathrm{d}u C' \bar{\beta}_{\perp}$ by U_1 and the limit of $\bar{\beta}_{\perp}' S_{10}$ by U_2, the sum of the solutions of $\det[\bar{\beta}_{\perp}' \{S(\lambda) - S(\lambda)\beta[\beta' S(\lambda)\beta]^{-1}\beta' S(\lambda)\}\bar{\beta}_{\perp}] = 0$ multiplied by T converges weakly toward the trace of $N_0^{1/2} U_2' U_1^{-1} U_2 N_0^{1/2}$.

Remark 5. The value of this result is limited by the fact that the asymptotic distribution is dependent of unknown parameters. A parametric bootstrap is one possibility to deal with this problem. A reasonable way to proceed may be first to estimate the parameters as described in Section 3 and compute the centered residuals $\hat{\epsilon}_t,\ t = k+1, \dots, T$. The bootstrap generated observations can then found using the representation described in Proposition 1 after resampling the centered residuals. Since there may be roots of the determinant of the estimated characteristic polynomial inside and/or outside the unit circle the recursions in Corollary 2 must be used to generate the stationary part. This parallels the procedure used to generate the Monte Carlo simulations in Section 5 where the point is discussed in more detail.

5. NUMERICAL RESULTS

To get an impression of the finite sample properties, consistency and asymptotic normality of the estimators we present some Monte Carlo simulations. Additional results can be found in the supporting information. The time series were generated from model (3) using the representation from Proposition 1.

Table I. Empirical mean and standard deviations of simulated estimates of β, VAR(1)

		$T=100$		$T=200$		$T=500$		$T=1000$	
	True	Mean	SD	Mean	SD	Mean	SD	Mean	SD
$\nu = 6,\ \beta_{31}$	0.0	−0.146	3.376	0.000	0.332	−0.002	0.056	0.00	0.015
$\nu = 6,\ \beta_{32}$	0.0	−0.043	13.642	−0.174	5.051	0.015	0.617	−0.01	0.358
$\nu = 20,\ \beta_{31}$	0.0	0.020	0.776	−0.016	0.514	0.001	0.054	−0.001	0.036
$\nu = 20,\ \beta_{32}$	0.0	0.136	4.697	−0.055	3.732	−0.013	0.628	0.013	0.638

Table II. Empirical mean and standard deviations of simulated estimates, VAR(1)

		$T=100$		$T=200$		$T=500$		$T=1000$	
	True	Mean	SD	Mean	SD	Mean	SD	Mean	SD
α_{11}	−0.5	−0.484	0.101	−0.460	0.116	−0.489	0.043	−0.494	0.026
α_{21}	0.0	0.011	0.129	0.002	0.125	0.003	0.064	0.001	0.044
α_{31}	0.0	−0.002	0.088	0.004	0.064	0.002	0.037	0.001	0.027
α_{12}	0.0	0.114	0.929	0.012	1.142	0.024	0.629	0.004	0.427
α_{22}	2.0	2.320	1.061	2.141	0.784	2.060	0.381	2.026	0.249
α_{32}	0.0	−0.037	0.909	−0.013	0.616	0.010	0.366	−0.002	0.255
σ_{11}	1.0	1.048	0.128	1.072	0.159	1.023	0.060	1.011	0.035
σ_{12}	0.0	0.014	0.357	0.006	0.385	0.009	0.208	0.002	0.145
σ_{22}	1.0	1.056	0.340	0.988	0.239	0.998	0.131	0.999	0.087
σ_{13}	0.0	−0.016	0.145	−0.002	0.113	−0.002	0.054	−0.003	0.037
σ_{23}	0.0	−0.008	0.310	−0.002	0.207	0.002	0.126	−0.002	0.090
σ_{33}	1.0	0.989	0.091	0.992	0.065	0.998	0.041	0.999	0.030
ν	6.0	9.152	18.012	6.583	1.966	6.197	1.009	6.112	0.656

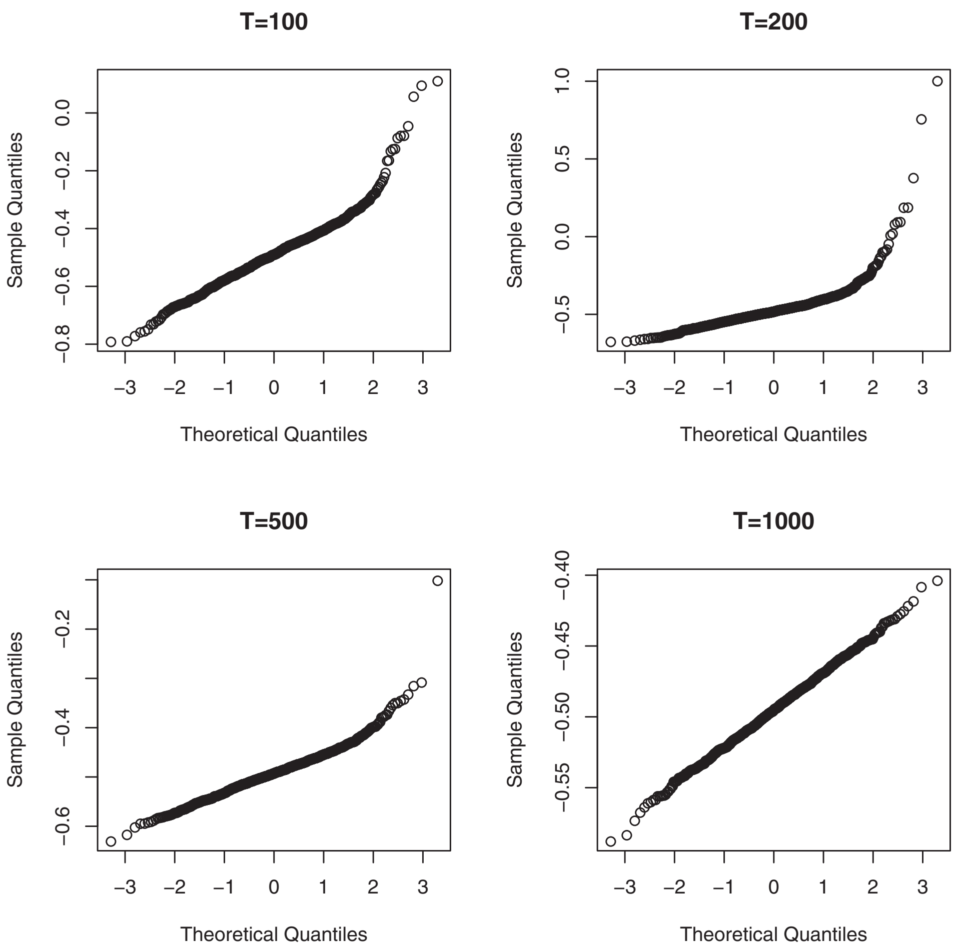

Figure 1. Estimates of $\alpha_{11} = -0.5$: QQ-plots for simulated values

The errors were p-variate elliptical t-distributed with expectation zero, ν degrees of freedom and parameters $\Sigma = \{\sigma_{ij}\}$ where Σ is a positive definite matrix with square root $\Sigma^{1/2}$. The density is

$$f_\epsilon(x) = \det(\Sigma)^{-1/2} \frac{\nu^{\nu/2}\Gamma((\nu+p)/2)}{\pi^{p/2}\Gamma(\nu/2)} (\nu + x'\Sigma^{-1}x)^{-(\nu+p)/2}. \tag{18}$$

This is the density of a random variable X which can be represented as $X = Z^{-1/2}(\nu\Sigma)^{1/2}Y$ where Y and Z are independent. The p-dimensional variable Y is multivariate normal with expectation 0 and covariance matrix I_p and Z is χ^2-distributed with ν degrees of freedom, see for example, Muirhead (1982).

The simulations were from a VAR(1) model of the form (3) with dimension $p = 3$, rank $r = 2$, $\alpha = \begin{pmatrix} -0.5 & 0 & 0 \\ 0 & 2.0 & 0 \end{pmatrix}'$, $\beta = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}'$, $\Sigma = I_3$ and $\nu = 6$ or $\nu = 20$. The absolute values of the solutions to $\det A(z) = 0$ are 3.0, 1.0 and 0.5.

For non-Gaussian errors the stationary distribution does not depend only on the two first moments of the error distribution. Therefore the recursions described in Corollary 2 with a burn-in period of 20 for each of the initial distributions of the forward- and backward recursions were used to simulate the stationary part of the process.

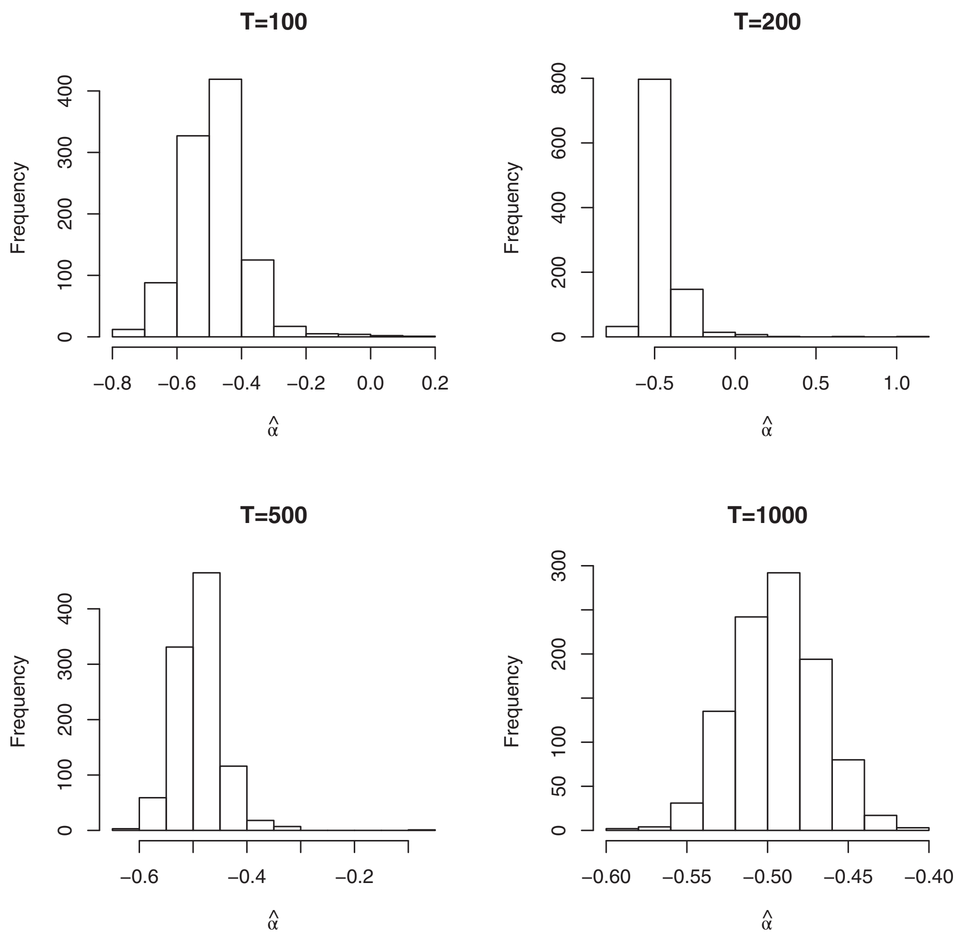

Figure 2. Estimates of $\alpha_{11} = -0.5$: histograms for simulated values

Table I illustrates the consistency of the estimator of β based on 1000 replications, as the length of the simulated series increases, solving the generalized eigenvalue problem discussed in Section 3.2. As one can see the distributions get more concentrated about the true values $(0, 0)'$ with the exception for β_{32} when $\nu = 20$ and T passes from 500 to 1000. The value is not significantly different from 0, however.

For maximizing the likelihood for β fixed, as explained in Section 3.2, the default Nelder–Mead option in the R-package `optim`, R Core Team (2016), was used supplemented by employing the BFGS option with numerical calculation of the gradients. The iterations were started with the values used for the simulations.

In Table II the simulations for estimating the parameters are summarized. For the case $T = 100$ only the Nelder–Mead step was possible. The QQ-plots and histograms in Figures 1 and 2 show in more detail the convergence of the estimates of the parameter α_{11}. The rather slow convergence toward normality may be related to problems in locating the maximum value of the likelihood which is more pronounced for small values of T. For $T = 100$ numerical derivatives of the likelihood were not always returned.

6. CONCLUSION

A main result of this article is a representation theorem of a p-dimensional autoregressive time series X_t where the roots of the determinant of the characteristic polynomial can be outside and inside the unit circle or equal to 1.

Under quite general conditions there exists a $p \times r$ matrix β such that the time series $\beta' X_t$ is stationary. The matrix β can be consistently estimated by the common generalized eigenvalue procedure. For fixed values of β and under some regularity conditions the maximum likelihood estimates of the remaining parameters are asymptotically Gaussian distributed. Also the asymptotic distribution of the trace test was discussed.

An open problem that remains is to find the asymptotic distribution of a consistent estimator for the parameter β. Also, further investigation of tests for hypotheses on this parameter is necessary. To find a procedure for determining the rank of $\Pi = \alpha\beta'$ is another challenge.

ACKNOWLEDGEMENTS

The article was presented at a seminar at University of California, San Diego in April 2017 and at the University of Oslo in June 2019. The author thanks the participants, in particular John Dagsvik, for valuable comments. Also comments from the referees and the editor were helpful.

DATA AVAILABILITY STATEMENT

R-code for the simulations in Section 5 is available from the author on request.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

REFERENCES

Andrews B, Davis RA, Breidt FJ. 2006. Maximum likelihood estimation for all-pass time series models. *Journal of Multivariate Analysis* **97**: 1638–1659.

Breidt FJ, Davis RA, Lii K-S, Rosenblatt M. 1991. Maximum likelihood estimation for noncausal autoregressive processes. *Journal of Multivariate Analysis* **36**: 175–198.

Brockwell PJ, Davis RA. 1991. *Time Series: Theory and Methods*, 2nd. ed. New York: Springer.

Cavaliere G, Rahbek A, Taylor AMR. 2012. Bootstrap determination of the co-integration rank in vector autoregressive models. *Econometrica* **80**: 1721–1740.

Cavaliere G, Nielsen HB, Rahbek A. 2020. Bootstrapping noncausal autoregressions: with applications to explosive bubble modeling. *Journal of Business & Econometric Statistics* **38**: 55–67.

Cubadda G, Hecq A, Telg S. 2019. Detecting co-movements in noncausal time series. *Oxford Bulletin of Economics and Statistics* **81**: 697–715.

Davis RA, Song L. 2020. Noncausal vector AR processes with application to economic time series. *Journal of Econometrics* **216**: 246–267.

Engle RF, Granger CWJ. 1987. Co-integration and error correction: representation, estimation and testing. *Econometrica* **55**: 251–276.

Engsted T, Nielsen B. 2012. Testing for rational bubbles in a coexplosive vector autoregression. *Econometrics Journal* **15**: 226–254.

Hannan EJ. 1970. *Multiple Time Series*. New York: Wiley.

Hannan EJ, Deistler M. 1988. *The Statistical Theory of Linear Systems*. New York: Wiley.

Hansen PR. 2005. Granger's representation theorem: a closed-form expression for $I(1)$ processes. *Econometrics Journal* **8**: 23–38.

Hendry DF. 1995. *Dynamic Econometrics*. Oxford: Oxford University Press.

Horn RA, Johnson CR. 2013. *Matrix Analysis*, 2nd ed. Cambridge: Cambridge University Press.

Gouriéroux C, Jasiak J. 2017. Noncausal vector autoregressive process: representation, identification and semi-parametric estimation. *Journal of Econometrics* **200**: 118–134.

Gouriéroux C, Zakoian J-M. 2017. Local explosion modeling by non-causal process. *Journal of the Royal Statistical Society B* **79**: 737–756.

Johansen S. 1995. *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*. Oxford: Oxford University Press.

Johansen S. 2009. Representation of cointegrated autoregressive processes with application to fractional processes. *Econometric Reviews* **28**: 121–145.

Johansen S, Juselius K. 2014. An asymptotic invariance property of the common trends under linear transformation of the data. *Journal of Econometrics* **178**: 310–315.

DOI: 10.1111/jtsa.12607 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd.

Lanne M, Saikkonen P. 2013. Noncausal vector autoregression. *Econometric Theory* **29**: 447–481.
Lütkepohl H. 2005. *New Introduction to Multiple Time Series Analysis*. Berlin: Springer.
Muirhead RJ. 1982. *Aspects of Multivariate Statistical Theory*. New York: Wiley.
Nielsen B. 2010. Analysis of coexplosive processes. *Econometric Theory* **26**: 882–915.
R Core Team. 2016. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
Rosenblatt M. 2000. *Gaussian and Non-Gaussian Linear Time Series and Random Fields*. New York: Springer.
Saikkonen P, Sandberg R. 2016. Testing for unit root in noncausal autoregressive models. *Journal of Time Series Analysis* **37**: 99–125.
Swensen AR. 2006. Bootstrap algorithms for testing and determining the cointegration rank in VAR models. *Econometrica* **74**: 1699–1714. Corrigendum. *Econometrica* **77**: 1703–1704.
van der Wart AW. 1998. *Asymptotic Statistics*. Cambridge: Cambridge University Press.

APPENDIX

A. PROOF OF THE SUFFICIENT PART OF PROPOSITION 1

By defining a suitable companion matrix the model (3) can be written compactly, see for example, Hansen (2005). Let the $pk \times l,\ l = p(k-1) + r$, matrices α^* and β^* be defined as

$$\alpha^* = \begin{pmatrix} \alpha & \Gamma_1 & \cdots & \Gamma_{k-1} \\ 0 & I & & 0 \\ \vdots & \ddots & & \\ 0 & & & I \end{pmatrix}, \beta^* = \begin{pmatrix} \beta & I & 0 & \cdots & 0 \\ 0 & -I & I & & 0 \\ \vdots & & \ddots & & \\ & & & & I \\ 0 & \cdots & & & -I \end{pmatrix}.$$

Then, the model (3) can be expressed as

$$\Delta X_t^* = \alpha^* \beta^{*\prime} X_{t-1}^* + \Phi_t^* + \epsilon_t^*,\ t = k+1, \ldots \tag{A.1}$$

where $X_t^* = (X_t', \ldots, X_{t-k+1}')'$, $\epsilon_t^* = (\epsilon_t', 0, \ldots, 0)'$ and $\Phi_t^* = ((\Phi D_t)', 0, \ldots, 0)'$.

Multiplying both sides of (A.1) with $\beta^{*\prime}$ and rearranging yields

$$\beta^{*\prime} X_t^* = (I + \beta^{*\prime}\alpha^*)\beta^{*\prime} X_{t-1}^* + \beta^{*\prime}(\epsilon_t^* + \Phi_t^*). \tag{A.2}$$

Lemma 1. Under Assumption 1 there exist non-singular real matrices M, G_1 and G_2 so that

$$I + \beta^{*\prime}\alpha^* = MGM^{-1} = M \begin{pmatrix} G_1 & 0 & 0 \\ 0 & G_2 & 0 \\ 0 & 0 & 0 \end{pmatrix} M^{-1}$$

where all the eigenvalues of G_1 have modulus less than 1, all the eigenvalues of G_2 have modulus larger than 1 and the lower right block is present when $I + \beta^{*\prime}\alpha^*$ is singular.

Proof. As shown in Lemma A.4 in Hansen (2005) the matrix $(\beta^*, \alpha_\perp^*)$ is non-singular. It then follows from his Lemma A.1 that the matrix $I + \beta^{*\prime}\alpha^*$ does not have 1 as an eigenvalue. Also, it is argued in the proof of Lemma A.2 that an eigenvalue, $\lambda \neq 0$, of $I + \beta^{*\prime}\alpha^*$ satisfies $\det A(1/\lambda) = 0$. By Assumption 1 (i) there can be no eigenvalue of modulus 1 of $I + \beta^{*\prime}\alpha^*$. Hence all the eigenvalues of $I + \beta^{*\prime}\alpha^*$ have either modulus less than 1 or larger than 1.

One possible way to carry out the construction of M is to appeal to Theorem 3.4.1.5 in Horn and Johnson (2013) and define G as the real Jordan canonical form and let the lower right block correspond to the eigenvalues equal to zero. The matrices G_1 and G_2 are found by rearranging the blocks in G. The real matrix M can then be found by appealing to Theorem 1.3.29 in Horn and Johnson (2013). □

From Lemma 1 it follows that the autoregressive scheme (A.2) can be written

$$M^{-1}\beta^{*\prime}X_t^* = \begin{pmatrix} G_1 & 0 & 0 \\ 0 & G_2 & 0 \\ 0 & 0 & 0 \end{pmatrix} M^{-1}\beta^{*\prime}X_{t-1}^* + M^{-1}\beta^{*\prime}(\epsilon_t^* + \Phi_t^*). \tag{A.3}$$

Since all diagonal elements in G_1 and G_2 have absolute values strictly larger or smaller than one, there exists a stationary solution of (A.2) of the form

$$\beta^{*\prime}X_t^* = M \sum_{s=-\infty}^{\infty} C_s^* M^{-1}\beta^{*\prime}(\epsilon_{t-s}^* + \Phi_{t-s}^*)$$

where by stationary we mean, as in Hansen (2005), that $\beta^{*\prime}X_t^* - E(\beta^{*\prime}X_t^*)$ is stationary. In Hannan and Deistler (1988, p. 12) and in Davis and Song (2020) this is explained in more detail.

Using that $\alpha_\perp^* = (\alpha_\perp', -\alpha_\perp'\Gamma_1, \dots, -\alpha_\perp'\Gamma_{k-1})'$ yields $\alpha_\perp^{*\prime}\Delta X_s^* = \alpha_\perp^{*\prime}(\epsilon_s^* + \Phi_s^*)$, $s = k+1, \dots, t$ and by summing $\alpha_\perp^{*\prime}(X_t^* - X_k^*) = \sum_{s=k+1}^{t} \alpha_\perp^{*\prime}(\epsilon_s^* + \Phi_s^*)$. Stacking $\beta^{*\prime}X_t^*$ and $\alpha_\perp^{*\prime}X_t^*$ yields the representation

$$X_t^* = (\beta^*, \alpha_\perp^*)'^{-1} \begin{pmatrix} M\sum_{s=-\infty}^{\infty} C_s^* M^{-1}\beta^{*\prime}(\epsilon_{t-s}^* + \Phi_{t-s}^*) \\ \sum_{s=k+1}^{t} \alpha_\perp^{*\prime}(\epsilon_s^* + \Phi_s^*) + \alpha_\perp^{*\prime}X_k^* \end{pmatrix}.$$

Hansen (2005) in the proof of his Theorem 1 showed that the upper $p \times pk$ sub-matrix of $(\beta^*, \alpha_\perp^*)'^{-1}$ can be written as $(F, C\bar{\alpha}_\perp)$ with $F = ((I - C\Gamma)\bar{\beta}, -C\Gamma_1^*, \dots, -C\Gamma_{k-1}^*)$ and $\Gamma_i^* = \Gamma_i + \dots + \Gamma_{k-1}$. Thus, with initial value $A = C\bar{\alpha}_\perp\alpha_\perp^{*\prime}X_k^* = \beta_\perp(\alpha_\perp'\Gamma\beta_\perp)^{-1}\alpha_\perp^{*\prime}X_k^* = C(X_k - \Gamma_1 X_{k-1} - \dots - \Gamma_{k-1}X_1)$,

$$\begin{aligned} X_t &= FM\sum_{s=-\infty}^{\infty} C_s^* M^{-1}\beta^{*\prime}(\epsilon_{t-s}^* + \Phi_{t-s}^*) + C\sum_{s=k+1}^{t}(\epsilon_s + \Phi D_s) + A \\ &= FM\sum_{s=-\infty}^{\infty} C_s^* M^{-1}B(\epsilon_{t-s} + \Phi D_{t-s}) + C\sum_{s=k+1}^{t}(\epsilon_s + \Phi D_s) + A, \ t = k+1, \dots \end{aligned}$$

with $B' = (\beta, I, 0, \dots, 0)$ so $\beta^{*\prime}(\epsilon_t^* + \Phi_t^*) = B(\epsilon_t + \Phi D_t)$.

B. PROOF OF COROLLARIES OF PROPOSITION 1

Proof of Corollary 1. To prove Corollary 1, define the power series $C_1(z) = \sum_{i=0}^{\infty} C_{1i}z^i$ and $C_2(z) = \sum_{i=1}^{\infty} C_{2i}z^i$ where C_{1i} and C_{2i} are the matrices consisting of the first p columns of $FM\tilde{C}_i^* M^{-1}$ and $FM\tilde{C}_{-i}^* M^{-1}$ respectively.

For a vector $x = (x_1, \dots, x_n)'$, consider the norm $\|x\|_\infty = \max_{1\le i\le n}|x_i|$ and for an $m \times n$ matrix $D = \{d_{ij}\}$ let $\|D\|_\infty$ be the induced norm, which equals $\max_{1\le i\le m}\sum_{j=1}^{n}|d_{ij}|$. For an $m \times n$ matrix A and $n \times o$ matrix B, $\|AB\|_\infty \le \|A\|_\infty\|B\|_\infty$. Therefore, because all the eigenvalues of the quadratic matrices $\tilde{C}_i^*$ and $\tilde{C}_{-i}^*$ have modulus strictly less than one, the elements in $C_1(z)$ and $C_2(z)$ must converge in a disk with radius $1 + \delta$ where $\delta > 0$, see Corollary A.2 in Johansen (1995). Taking $C(z) = C_1(z) + C_2(1/z)$ concludes the proof. □

Proof of Corollary 2. Consider first the variable $\tilde{X}_{1,t}$. Then for $t = k+1, \dots, T$

$$\tilde{X}_{1,t} = \sum_{s=0}^{\infty} \begin{pmatrix} G_1^s & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} M^{-1}B\epsilon_{t-s}$$

$$= \begin{pmatrix} I & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} M^{-1}B\epsilon_t + \sum_{s=1}^{\infty} \begin{pmatrix} G_1^s & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} M^{-1}B\epsilon_{t-s}$$

$$= \sum_{s=0}^{\infty} \begin{pmatrix} G_1^{s+1} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} M^{-1}B\epsilon_{t-1-s} + \begin{pmatrix} I & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} M^{-1}B\epsilon_t$$

$$= \begin{pmatrix} G_1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tilde{X}_{1,t-1} + \begin{pmatrix} I & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} M^{-1}B\epsilon_t.$$

Similarly for variable $\tilde{X}_{2,t}$. Then

$$\tilde{X}_{2,t} = -\sum_{s=-\infty}^{-1} \begin{pmatrix} 0 & 0 & 0 \\ 0 & G_2^s & 0 \\ 0 & 0 & 0 \end{pmatrix} M^{-1}B\epsilon_{t-s}$$

$$= -\begin{pmatrix} 0 & 0 & 0 \\ 0 & G_2^{-1} & 0 \\ 0 & 0 & 0 \end{pmatrix} M^{-1}B\epsilon_{t+1} - \sum_{s=-\infty}^{-2} \begin{pmatrix} 0 & 0 & 0 \\ 0 & G_2^s & 0 \\ 0 & 0 & 0 \end{pmatrix} M^{-1}B\epsilon_{t-s}$$

$$= -\sum_{s=-\infty}^{-1} \begin{pmatrix} 0 & 0 & 0 \\ 0 & G_2^{s-1} & 0 \\ 0 & 0 & 0 \end{pmatrix} M^{-1}B\epsilon_{t+1-s} - \begin{pmatrix} 0 & 0 & 0 \\ 0 & G_2^{-1} & 0 \\ 0 & 0 & 0 \end{pmatrix} M^{-1}B\epsilon_{t+1}$$

$$= \begin{pmatrix} 0 & 0 & 0 \\ 0 & G_2^{-1} & 0 \\ 0 & 0 & 0 \end{pmatrix} \tilde{X}_{2,t+1} - \begin{pmatrix} 0 & 0 & 0 \\ 0 & G_2^{-1} & 0 \\ 0 & 0 & 0 \end{pmatrix} M^{-1}B\epsilon_{t+1},\ t = T, \dots, k+1.$$

□

C. PROOF OF PROPOSITION 2

Define $\varepsilon_t = (\varepsilon_t^{1\prime}, \varepsilon_t^{2\prime})' = M_1^{-1}\beta^{*\prime}\epsilon_t^*$. Then

Lemma 2. (i) There exists a non-singular matrix T_0 with $\det T_0 = (-\det J_2^{-1})^{T-k}$ such that $(V_k', \dots, V_T')' = T_0(V_k^{1\prime}, \varepsilon_{k+1}', \dots, \varepsilon_T', V_T^{2\prime})'$.
(ii) There exists a non-singular matrix H_1 with $\det H_1 = (detM_1)^{-(T-k)}$ such that $(V_k^{1\prime}, \varepsilon_{k+1}', \dots, \varepsilon_T', V_T^{2\prime})' = H_1(V_k^{1\prime}, \epsilon_{k+1}^{*\prime}\beta^* \dots, \epsilon_T^{*\prime}\beta^*, V_T^{2\prime})'$.

Proof. (i) Solving $V_t^1 = J_1 V_{t-1}^1 + \varepsilon_t^1$ backward and arranging on a matrix form

$$\begin{pmatrix} V_k^1 \\ V_{k+1}^1 \\ V_{k+2}^1 \\ \vdots \\ V_T^1 \end{pmatrix} = \begin{pmatrix} I & 0 & 0 & \cdots & 0 \\ J_1 & I & 0 & \cdots & 0 \\ J_1^2 & J_1 & 0 & \cdots & 0 \\ \vdots & & & & \vdots \\ J_1^{T-k} & J_1^{T-k-1} & J_1^{T-k-2} & \cdots & I \end{pmatrix} \begin{pmatrix} V_k^1 \\ \varepsilon_{k+1}^1 \\ \varepsilon_{k+2}^1 \\ \vdots \\ \varepsilon_T^1 \end{pmatrix} = T_1 \begin{pmatrix} V_k^1 \\ \varepsilon_{k+1}^1 \\ \varepsilon_{k+2}^1 \\ \vdots \\ \varepsilon_T^1 \end{pmatrix}.$$

J. Time Ser. Anal. **43**: 178–196 (2022)
Journal of Time Series Analysis published by John Wiley & Sons Ltd. DOI: 10.1111/jtsa.12607

Similarly solving $V_t^2 = J_2 V_{t-1}^2 + \varepsilon_t^2$ forward and arranging

$$\begin{pmatrix} V_k^2 \\ V_{k+1}^2 \\ V_{k+2}^2 \\ \vdots \\ V_T^2 \end{pmatrix} = \begin{pmatrix} -J_2^{-1} & -J_2^{-2} & -J_2^{-3} & \cdots & -J_2^{-(T-k)} & J_2^{-(T-k)} \\ 0 & -J_2^{-1} & -J_2^{-2} & \cdots & -J_2^{-(T-(k+1))} & J_2^{-(T-(k+1))} \\ \vdots & & & & & \vdots \\ 0 & 0 & 0 & \cdots & -J_2^{-1} & J_2^{-1} \\ 0 & 0 & 0 & \cdots & 0 & I \end{pmatrix} \begin{pmatrix} \varepsilon_{k+1}^2 \\ \varepsilon_{k+2}^2 \\ \vdots \\ \varepsilon_T^2 \\ V_T^2 \end{pmatrix} = T_2 \begin{pmatrix} \varepsilon_{k+1}^2 \\ \varepsilon_{k+2}^2 \\ \vdots \\ \varepsilon_T^2 \\ V_T^2 \end{pmatrix}.$$

Thus for suitable permutation matrices P_1 and P_2

$$\begin{pmatrix} V_k \\ \vdots \\ V_T \end{pmatrix} = P_1 \begin{pmatrix} V_k^1 \\ V_{k+1}^1 \\ \vdots \\ V_T^1 \\ V_k^2 \\ V_{k+1}^2 \\ \vdots \\ V_T^2 \end{pmatrix} = P_1 \begin{pmatrix} T_1 & 0 \\ 0 & T_2 \end{pmatrix} \begin{pmatrix} V_k^1 \\ \varepsilon_{k+1}^1 \\ \varepsilon_{k+2}^1 \\ \vdots \\ \varepsilon_T^1 \\ \varepsilon_{k+1}^2 \\ \varepsilon_{k+2}^2 \\ \vdots \\ \varepsilon_T^2 \\ V_T^2 \end{pmatrix} = P_1 \begin{pmatrix} T_1 & 0 \\ 0 & T_2 \end{pmatrix} P_2 \begin{pmatrix} V_k^1 \\ \varepsilon_{k+1} \\ \varepsilon_{k+2} \\ \vdots \\ \varepsilon_T \\ V_T^2 \end{pmatrix}.$$

Let

$$T_0 = P_1 \begin{pmatrix} T_1 & 0 \\ 0 & T_2 \end{pmatrix} P_2.$$

(ii) With

$$H_1 = \begin{pmatrix} I & 0 & 0 & \cdots & 0 \\ 0 & M_1^{-1} & & \cdots & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & 0 \cdots & M_1^{-1} & 0 \\ 0 & 0 & 0 \cdots & 0 & I \end{pmatrix}.$$

the claim follows from the definition of $\varepsilon_{k+1}, \ldots, \varepsilon_T$. □

Proof of Proposition 2. Since $\Delta X_t^* = X_t^* - X_{t-1}^* = \alpha^* \beta^{*\prime} X_{t-1}^* + \epsilon_t^*,\ t = k+1, \ldots, T$ and $(\beta^*, \alpha_\perp^*)'$ is invertible, see Lemma A.1 in Hansen (2005),

$$\begin{pmatrix} I & 0 & \cdots & 0 \\ 0 & (\beta^*, \alpha_\perp^*)' & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \cdots & (\beta^*, \alpha_\perp^*)' \end{pmatrix} \begin{pmatrix} V_k \\ \Delta X_{k+1}^* \\ \vdots \\ \Delta X_T^* \end{pmatrix} = \begin{pmatrix} V_k \\ \beta^{*\prime} X_{k+1}^* - \beta^{*\prime} X_k^* \\ \alpha_\perp^{*\prime} \alpha^* \beta^{*\prime} X_k^* + \alpha_\perp^{*\prime} \epsilon_{k+1}^* \\ \vdots \\ \beta^{*\prime} X_T^* - \beta^{*\prime} X_{T-1}^* \\ \alpha_\perp^{*\prime} \alpha^* \beta^{*\prime} X_{T-1}^* + \alpha_\perp^{*\prime} \epsilon_T^* \end{pmatrix}$$

DOI: 10.1111/jtsa.12607

$$
= \begin{pmatrix}
I & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\
-M_1 & M_1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\
0 & 0 & I & 0 & 0 & \ldots & 0 & 0 & 0 \\
0 & -M_1 & 0 & M_1 & 0 & \cdots & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & I & \ldots & 0 & 0 & 0 \\
\vdots & & & & & & & & \vdots \\
 & & & & & -M_1 & 0 & M_1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & I
\end{pmatrix}
\begin{pmatrix} V_k \\ V_{k+1} \\ \alpha_\perp^{*\prime}\epsilon_{k+1}^* \\ \vdots \\ V_T \\ \alpha_\perp^{*\prime}\epsilon_T^* \end{pmatrix}
$$

$$
= H_0 \begin{pmatrix} V_k \\ V_{k+1} \\ \alpha_\perp^{*\prime}\epsilon_{k+1}^* \\ \vdots \\ V_T \\ \alpha_\perp^{*\prime}\epsilon_T^* \end{pmatrix} = H_0 P_3 \begin{pmatrix} V_k \\ \vdots \\ V_T \\ \alpha_\perp^{*\prime}\epsilon_{k+1} \\ \vdots \\ \alpha_\perp^{*\prime}\epsilon_T \end{pmatrix}
$$

where P_3 is a permutation matrix. Therefore by Lemma 2 and using $V_k = M_1^{-1}\beta^{*\prime}X_k$

$$
\begin{pmatrix}
M_1^{-1} & 0 & \cdots & 0 \\
0 & (\beta^*, \alpha_\perp^*)' & \cdots & 0 \\
\vdots & & & \vdots \\
0 & 0 & \cdots & (\beta^*, \alpha_\perp^*)'
\end{pmatrix}
\begin{pmatrix} \beta^{*\prime}X_k^* \\ \Delta X_{k+1}^* \\ \vdots \\ \Delta X_T^* \end{pmatrix}
= H_0 P_3 \begin{pmatrix} T_0 H_1 & 0 \\ 0 & I \end{pmatrix}
\begin{pmatrix} V_k^1 \\ \beta^{*\prime}\epsilon_{k+1}^* \\ \vdots \\ \beta^{*\prime}\epsilon_T^* \\ V_T^2 \\ \alpha_\perp^{*\prime}\epsilon_{k+1}^* \\ \vdots \\ \alpha_\perp^{*\prime}\epsilon_T^* \end{pmatrix}
$$

$$
= H_0 P_3 \begin{pmatrix} T_0 H_1 & 0 \\ 0 & I \end{pmatrix} P_4
\begin{pmatrix}
I & 0 & \cdots & 0 & 0 \\
0 & (\beta^*, \alpha_\perp^*)' & \cdots & 0 & 0 \\
\vdots & & & & \vdots \\
0 & 0 & \cdots & (\beta^*, \alpha_\perp^*)' & 0 \\
0 & 0 & \cdots & 0 & I
\end{pmatrix}
\begin{pmatrix} V_k^1 \\ \epsilon_{k+1}^* \\ \vdots \\ \epsilon_T^* \\ V_T^2 \end{pmatrix}
$$

for another permutation matrix P_4. Here $\det H_0 = (\det M_1)^{T-k}$, $\det H_1 = (\det M_1)^{-(T-k)}$ and $\det T_0 = (-\det J_2^{-1})^{T-k}$. □