

Article

Information-Criterion-Based Lag Length Selection in Vector Autoregressive Approximations for I(2) Processes

Dietmar Bauer 

Department of Business Administration and Economics, Bielefeld University, Universitätsstrasse 25, D-33615 Bielefeld, Germany; dietmar.bauer@uni-bielefeld.de

Abstract: When using vector autoregressive (VAR) models for approximating time series, a key step is the selection of the lag length. Often this is performed using information criteria, even if a theoretical justification is lacking in some cases. For stationary processes, the asymptotic properties of the corresponding estimators are well documented in great generality in the book Hannan and Deistler (1988). If the data-generating process is not a finite-order VAR, the selected lag length typically tends to infinity as a function of the sample size. For invertible vector autoregressive moving average (VARMA) processes, this typically happens roughly proportional to $\log T$. The same approach for lag length selection is also followed in practice for more general processes, for example, unit root processes. In the I(1) case, the literature suggests that the behavior is analogous to the stationary case. For I(2) processes, no such results are currently known. This note closes this gap, concluding that information-criteria-based lag length selection for I(2) processes indeed shows similar properties to in the stationary case.

Keywords: long VAR approximation; lag length selection; I(2) processes

JEL Classification: C13; C32



Citation: Bauer, Dietmar. 2023. Information-Criterion-Based Lag Length Selection in Vector Autoregressive Approximations for I(2) Processes. *Econometrics* 11: 11. <https://doi.org/10.3390/econometrics11020011>

Academic Editor: Tae-Hwy Lee

Received: 15 December 2022

Revised: 17 April 2023

Accepted: 18 April 2023

Published: 20 April 2023



Copyright: © 2023 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The workhorse in empirical modeling of economic time series is still constituted by the vector autoregressive (VAR) model. The VAR model explains the process $(y_t)_{t \in \mathbb{Z}}, y_t \in \mathbb{R}^p$, using the difference equation:

$$y_t = A_1 y_{t-1} + \dots + A_h y_{t-h} + u_t, \quad t \in \mathbb{Z} \quad (1)$$

where $(u_t)_{t \in \mathbb{Z}}$ is a white noise process. Under the stability assumption $|A(z)| \neq 0, |z| \leq 1$ for $A(z) = I_p - A_1 z - \dots - A_h z^h$, this equation has a unique stationary solution. Often, prior to VAR modeling, deterministic components are extracted such as a constant, a linear trend, and seasonal dummies.

Oftentimes, this model is seen as an approximation to the data-generating process. This happens, for instance, if the data are generated by a vector autoregressive moving average (VARMA) process, where $u_t = \varepsilon_t + B_1 \varepsilon_{t-1} + \dots + B_g \varepsilon_{t-g}$ for white noise $(\varepsilon_t)_{t \in \mathbb{Z}}$ with expectation zero and variance $\Sigma > 0$. The VARMA system $(A(z), B(z)), B(z) = I_p + B_1 z + \dots + B_g z^g$ (where the pair $(A(z), B(z))$ is left-coprime) is called invertible if $|B(z)| \neq 0, |z| \leq 1$. Under stability and invertibility, we obtain the VAR(∞) representation

$$y_t = \sum_{j=1}^{\infty} \Phi_j y_{t-j} + \varepsilon_t.$$

Under invertibility, we have

$$\|\Phi_j\| \leq C \rho_0^j \quad (2)$$

where $\rho_0^{-1} > 1$ typically is the smallest modulus of the solutions of $|B(z)| = 0$. Thus, the coefficients converge to zero, geometrically implying that a VAR approximation seems plausible.

When estimating a VAR model, one has to decide on the lag length, h . This choice of the lag length is typically based on adding a penalty term to the estimation accuracy as measured using the residual variance Σ_h for a range of potential lag lengths $0 \leq h \leq H_T$, resulting in so-called *information criteria*. Here, the residual variance Σ_h is estimated from data as follows:¹

$$\hat{\Sigma}_h = T^{-1} \sum_{t=h+1}^T \hat{\varepsilon}_t(h)\hat{\varepsilon}_t(h)', \quad \hat{\varepsilon}_t(h) = y_t - \hat{A}_1 y_{t-1} - \dots - \hat{A}_h y_{t-h}. \tag{3}$$

If no structural restrictions apply, the matrices $\hat{A}_1, \dots, \hat{A}_h$ are typically estimated using OLS or the Yule–Walker equations (see, for example, Hannan and Deistler 1988, p. 211).² In this paper, we only consider OLS estimation.

Information criteria then are defined as

$$IC(h; C_T) = \log \det \hat{\Sigma}_h + \frac{C_T h p^2}{T}. \tag{4}$$

Here, $h p^2$ denotes the number of parameters estimated in the conditional mean equation, and C_T is a penalty factor. The most prominent choices are AIC ($C_T = 2$) and BIC ($C_T = \log T$). The minimal integer \hat{h} minimizing IC is then the selected lag length.³

For the case of stationary invertible VARMA processes, the asymptotic properties are well documented in section 6.6 of HD. Under appropriate assumptions on the noise (see Theorem 6.6.3 of HD) and upper bounds H_T for the lag length, for \hat{h}_{BIC} estimated using BIC, we have $\hat{h}_{BIC}/l(T) = 1$ a.s. (Theorem 6.6.3 of HD), and the same limit holds for \hat{h}_{AIC} in probability (Theorem 6.6.4 of HD), where $l(T) = \log T / (-2 \log \rho_0)$. Consequently, $\rho_0^{l(T)} = \exp(-\log \rho_0 \log T / (2 \log \rho_0)) = T^{-1/2}$. Thus, in this case, asymptotically, the approximation error (2) is of the order $T^{-1/2}$.

Moreover, Theorem 7.4.7 of HD states that in a more general setting, if $(y_t)_{t \in \mathbb{Z}}$ is not generated by a finite autoregression under the same assumptions on $(\varepsilon_t)_{t \in \mathbb{Z}}$, as above and where $\sum_{j=1}^{\infty} j^{1/2} \|\Phi_j\| < \infty$, we have

$$IC(h; C_T) = \log \det \hat{\Sigma}_T + \tilde{L}_T(h)(1 + o_P(1)), \quad \tilde{L}_T(h) = \frac{h p^2}{T} (C_T - 1) + \text{tr}[\Sigma^{-1}(\Sigma_h - \Sigma)], \tag{5}$$

uniformly in $0 \leq h \leq H_T = o(\sqrt{T/(\log T)})$, if $C_T > 1$. Here, $\hat{\Sigma}_T = \frac{1}{T} \sum_{t=1}^T \varepsilon_t \varepsilon_t'$ is independent of h . Now impose the following assumption on the sequence $\Sigma_h, h \in \mathbb{N}$:

Assumption 1. For the sequence $\Sigma_h, h \in \mathbb{N}$, there exists a twice continuously differentiable function $\theta(h)$ with second derivative $\theta''(h)$ such that $\lim_{h \rightarrow \infty} \{\text{tr}[\Sigma^{-1}(\Sigma_h - \Sigma)]\} / \theta(h) = 1$ and

$$\liminf_{h \rightarrow \infty} h^2 \theta''(h) / \theta(h) > 0.$$

Under this assumption HD show that the optimal h (for $IC(h; C_T)$) is close to a minimizer $l(T)$ of the deterministic function $L_T(h) = \frac{h p^2}{T} (C_T - 1) + \theta(h)$.

The limitation to stationary processes and autoregressive coefficients Φ_j declining fast enough excludes persistent processes such as unit root processes that are often encountered in economic time series. It is common practice, nevertheless, to select the lag length using information criteria also in case of highly persistent processes. For integrated processes of the I(1) kind, this is backed by some results in the literature, such as Ng and Perron (1995), stating that lag length selection for VARMA models has properties analogous to the stationary case. These results are generalized to a larger class of processes in Lütkepohl and Saikkonen (1999).⁴

However, in the literature, there are currently no results available for the I(2) case of doubly integrated processes. This class of processes was introduced by Granger and Lee (1989), who illustrated the main underlying idea using an example from inventory processes: if the demand for a product (a flow variable) is modeled as an I(1) process (which is realistic in a number of cases; see Granger and Lee 1989), then the stock of the product at the producers will be I(2), as stock variables sum up the corresponding flow variables. In a macroeconomic framework, I(2) processes have been investigated, for example, by Johansen and Juselius and coworkers, who argue that I(2) processes play a role if the model contains nominal quantities, as inflation often is found to be integrated. Since inflation is the change rate of prices, it follows that prices must be I(2) then (see, for example, Juselius 2006, chps. 16–18, or Johansen 1995, chp. 9).

While these models have been used in the literature, the corresponding inferential methodology focuses on the case of autoregressions with a finite lag length. If the data-generating process is more general, such models are often an approximation and must involve the selection of the lag length. In practice, this is performed using information criteria such as AIC or BIC also in this situation while a theoretical justification is missing.

This paper closes the following gap: we provide a general result for the asymptotic behavior of lag length selection using information criteria for I(2) processes, linking the performance for the I(2) processes to a related stationary process for which the results of HD can be applied. The proof of this result can be found in the appendix. The result is illustrated using a simulation study in Section 3.

2. Integrated Processes

The key to the results for integrated processes is the insight that the OLS residuals of (1) and—for example—of the vector error correction equation

$$\Delta y_t = \Pi y_{t-1} + \sum_{j=1}^{h-1} \Gamma_j \Delta y_{t-j} + u_t, \quad \Delta y_t = y_t - y_{t-1}, \quad (6)$$

are identical. Since the information criteria are a function of $\hat{\Sigma}_h$ (see (3)), they are invariant to linear transformations of the regressors as well as to transformations of the dependent variable by adding linear functions of the regressors.

If in (6) $\Pi = 0$, this defines a VAR($h - 1$) process $(\Delta y_t)_{t \in \mathbb{Z}}$ for which stationary theory applies under stability assumptions for $\Gamma(z) = I_p - \sum_{j=1}^{h-1} \Gamma_j z^j$. As in this case the estimation of Π is superconsistent, one would suspect that the inclusion of y_{t-1} as a regressor does not change the lag length selection substantially except for adding one lag to account for the differencing. A similar reasoning applies in the case of reduced rank of Π .

Paulsen (1984) shows that for finite-order autoregressive I(1) processes, indeed, the results for the stationary case can be extended. Lütkepohl and Saikkonen (1999) extend some of the results to VAR(∞) I(1) processes, rectifying earlier incorrect proofs in the literature. Bauer and Wagner (2004) provide theory for the case of multifrequency I(1) processes such as seasonally integrated processes by using the vector of seasons representation linking the multifrequency I(1) case to the I(1) case.

In more detail, let, in the I(1) case, $\Pi = \alpha\beta'$, where $\beta \in \mathbb{R}^{p \times r}$, $\beta'\beta = I_r$ and⁵ $\beta_\perp \in \mathbb{R}^{p \times (p-r)}$, $\beta'\beta_\perp = 0$, $\beta'_\perp\beta_\perp = I_{p-r}$ denote matrices such that $B = [\beta, \beta_\perp]$ is orthonormal. Let $\tilde{y}_t = B'y_t = (\tilde{y}'_{1,t}, \tilde{y}'_{2,t})'$, $\tilde{y}_{1,t} \in \mathbb{R}^r$. Then assume that

$$\begin{pmatrix} \tilde{y}_{1,t} \\ \Delta \tilde{y}_{2,t} \end{pmatrix} = u_t = \begin{pmatrix} u_{1,t} \\ u_{2,t} \end{pmatrix} = \tilde{c}_\bullet(L)\varepsilon_t, \quad (7)$$

where $|\tilde{c}_\bullet(z)| \neq 0, |z| \leq 1, \tilde{c}_\bullet(z) = \sum_{j=0}^\infty \tilde{c}_{\bullet,j}z^j$ and where it is assumed that $\sum_{j=0}^\infty j^{3/2}\|\tilde{c}_{\bullet,j}\| < \infty$. Under these assumptions, it follows that we obtain the VAR(∞) representation $\tilde{\Phi}(L)u_t = \varepsilon_t$ where $\tilde{\Phi}(z) = \tilde{c}_\bullet^{-1}(z)$, implying the VAR(∞) representation

$$\tilde{\Phi}(L)u_t = [\tilde{\Phi}(L)\text{diag}(I_r, \Delta I_{p-r})B']y_t = \varepsilon_t$$

for the I(1) process $(y_t)_{t \in \mathbb{Z}}$. Furthermore, the triangular representation (7) implies that the process $(\tilde{y}_{2,t})_{t \in \mathbb{Z}}$ is integrated and not cointegrated; compare, for example, Saikkonen and Lütkepohl (1996), sct. 2.

The corresponding VAR(h) approximation is given as $y_t = \sum_{j=1}^h A_j y_{t-j} + e_t(h)$. Pre-multiplying with B' and using that (for $h > 1$), the set of regressors $y_{t-1}, y_{t-2}, \dots, y_{t-h}$ (of dimension hp) can be linearly transformed into the set of regressors $u_{t-1}, \dots, u_{t-h+1}$ (of dimension $p(h-1)$) plus $\tilde{y}_{2,t-1} \in \mathbb{R}^{p-r}, u_{1,t-h} = \tilde{y}_{1,t-h} \in \mathbb{R}^r$, we obtain (for appropriate coefficient matrices)

$$\tilde{y}_t = \tilde{\Phi}_{1,2}\tilde{y}_{2,t-1} + \sum_{j=1}^{h-1} \tilde{\Phi}_j u_{t-j} + A_{h,1}u_{1,t-h} + \tilde{e}_t(h)$$

where $\tilde{e}_t(h) = B'e_t(h)$. Using (7) then leads to

$$u_t = \Phi_{1,2}\tilde{y}_{2,t-1} + \sum_{j=1}^{h-1} \tilde{\Phi}_j u_{t-j} + A_{h,1}u_{1,t-h} + \tilde{e}_t(h).$$

Here, all variables are stationary except for $\tilde{y}_{2,t-1}$, which is I(1) and not cointegrated by design. Consequently, $\Phi_{1,2} = 0$ has to hold (compare equation (A.2) in Saikkonen and Lütkepohl 1996). Omitting this regressor, the approximation almost equals a VAR(h) approximation of $(u_t)_{t \in \mathbb{Z}}$ except for the inclusion of $u_{1,t-h}$ instead of the full vector u_{t-h} . We have $\Sigma_{h-1} \geq \Sigma_{h-1}^u \geq \Sigma_h \geq \Sigma_h^u$ for each $h > 1$ where Σ_h^u denotes the residual variance in a VAR(h) approximation for $(u_t)_{t \in \mathbb{Z}}$. This shows that the residual variance of $(y_t)_{t \in \mathbb{Z}}$ and the one of $(u_t)_{t \in \mathbb{Z}}$ are closely related.

Thus, the essential step in linking the asymptotic properties of the lag length selection in the I(1) case (obtained from inclusion of $\tilde{y}_{2,t-1}$) to the ones in the stationary case (assuming that $\tilde{y}_{2,t-1}$ is left out, which is infeasible in practice as the matrix β is not known prior to estimation) lies in establishing that the inclusion of $\tilde{y}_{2,t-1}$ has negligible effects.

For the I(2) case, the same route can be taken. To this end, consider processes according to the following assumptions:

Assumption 2. Let $(y_t)_{t \in \mathbb{Z}}$ be an I(2) process (not generated by a finite-order autoregression) obtained as a solution to the equation (with deterministic values y_0, y_1 , and y_2)

$$\Delta^2 y_t = \alpha \beta' y_{t-1} + \Gamma \Delta y_{t-1} + \sum_{j=1}^\infty \Pi_j \Delta^2 y_{t-j} + \varepsilon_t, \quad t \in \mathbb{Z},$$

where

- $\alpha, \beta \in \mathbb{R}^{p \times r}, 0 \leq r < p$ are full column rank matrices;
- The function $\Pi(z) = (1-z)^2 I_p - \alpha \beta' z - \Gamma(1-z)z - \sum_{j=1}^\infty (1-z)^2 \Pi_j z^j$ (converging absolutely for $|z| < 1 + \delta, \delta > 0$) fulfills that $|\Pi(z)| = 0$ implies that $|z| > 1$ or $z = 1$;
- With $\beta_2 = \beta_\perp \eta_\perp, \alpha_2 = \alpha_\perp \xi_\perp$ (where $\alpha'_\perp \Gamma \beta_\perp = \xi \eta', \eta, \xi \in \mathbb{R}^{(p-r) \times s}$ are of full column rank $s < p - r$) the matrix

$$\alpha'_2 (I_p + \Gamma \beta (\beta' \beta)^{-1} (\alpha' \alpha)^{-1} \alpha' \Gamma - \sum_{j=1}^\infty \Pi_j) \beta_2 \tag{8}$$

is nonsingular.

Furthermore, the process $(\varepsilon_t)_{t \in \mathbb{Z}}$ is independent identically distributed (iid) with mean zero, variance $\Sigma > 0$, and with $\mathbb{E}(\varepsilon_{t,j}^4 \log_+ |\varepsilon_{t,j}|) < \infty, j = 1, \dots, p$.

From these assumptions, it follows that the process $(y_t)_{t \in \mathbb{Z}}$ is I(2) with cointegration and potentially multi-cointegration occurring. The structure of the process is best seen in the triangular representation of the process, which is equivalent to the one used above (see, for example, Li and Bauer 2020): assume that the stationary process $(u_t)_{t \in \mathbb{Z}}$ is generated according to $u_t = \tilde{c}_\bullet(L)\varepsilon_t$ (for suitable assumptions on $\tilde{c}_\bullet(z)$ including $|\tilde{c}_\bullet(1)| \neq 0$) and related to the process $(y_t)_{t \in \mathbb{Z}}$ via

$$\begin{pmatrix} \tilde{y}_{1,t} - A\Delta\tilde{y}_{3,t} \\ \Delta\tilde{y}_{2,t} \\ \Delta^2\tilde{y}_{3,t} \end{pmatrix} = u_t, \quad \tilde{y}_t = \begin{pmatrix} \tilde{y}_{1,t} \\ \tilde{y}_{2,t} \\ \tilde{y}_{3,t} \end{pmatrix} = \begin{pmatrix} \beta' \\ \beta'_1 \\ \beta'_2 \end{pmatrix} y_t,$$

where β, β_1 , and β_2 are as in Assumption 2. Then, clearly, $(\tilde{y}_{3,t})_{t \in \mathbb{Z}} = \beta'_2(y_t)_{t \in \mathbb{Z}}$ is I(2) and not cointegrated, $(\tilde{y}_{2,t})_{t \in \mathbb{Z}} = \beta'_1(y_t)_{t \in \mathbb{Z}}$ is I(1) and not cointegrated, and $(\tilde{y}_{1,t})_{t \in \mathbb{Z}} = \beta'(y_t)_{t \in \mathbb{Z}}$ is I(1) and cointegrates with $\Delta\tilde{y}_{3,t}$ to stationarity for nonzero A . If $A = 0$, it is stationary.⁶ The cointegrating relations $\tilde{y}_{1,t} - A\Delta\tilde{y}_{3,t}$ between an I(1) and a differenced I(2) process are termed *multi-cointegration* by Granger and Lee (1989).

It follows that (using $u_{3,t} = \Delta^2\tilde{y}_{3,t} = \Delta\tilde{y}_{3,t} - \Delta\tilde{y}_{3,t-1}$)

$$\begin{aligned} \Delta^2\tilde{y}_t &= \begin{pmatrix} \Delta^2\tilde{y}_{1,t} \\ \Delta^2\tilde{y}_{2,t} \\ \Delta^2\tilde{y}_{3,t} \end{pmatrix} = \begin{pmatrix} \tilde{y}_{1,t} - \tilde{y}_{1,t-1} - \Delta\tilde{y}_{1,t-1} \\ \Delta\tilde{y}_{2,t} - \Delta\tilde{y}_{2,t-1} \\ \Delta^2\tilde{y}_{3,t} \end{pmatrix} = u_t + \begin{pmatrix} A\Delta\tilde{y}_{3,t} - \tilde{y}_{1,t-1} - \Delta\tilde{y}_{1,t-1} \\ -\Delta\tilde{y}_{2,t-1} \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} -I & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tilde{y}_{t-1} + \begin{pmatrix} -I & 0 & A \\ 0 & -I & 0 \\ 0 & 0 & 0 \end{pmatrix} \Delta\tilde{y}_{t-1} + \underbrace{\begin{pmatrix} I & 0 & A \\ 0 & I & 0 \\ 0 & 0 & I \end{pmatrix}}_A u_t. \end{aligned}$$

Denoting $\mathcal{A}u_t = c_\bullet(L)\varepsilon_t$ and using $c_\bullet^{-1}(L) = c_\bullet^{-1}(1) + c_\bullet^*(1)\Delta + c_\bullet^{**}(L)\Delta^2$, we obtain from premultiplying the above equation with $c_\bullet^{-1}(L)$:

$$\begin{aligned} c_\bullet^{-1}(L)\Delta^2\tilde{y}_t &= \varepsilon_t - c_\bullet^{-1}(1) \begin{pmatrix} I \\ 0 \\ 0 \end{pmatrix} \beta' y_{t-1} \\ &\quad + \left\{ c_\bullet^*(1) \begin{pmatrix} -I \\ 0 \\ 0 \end{pmatrix} \beta' + c_\bullet^{-1}(1) \begin{pmatrix} -\beta' + A\beta'_2 \\ -\beta'_1 \\ 0 \end{pmatrix} \right\} \Delta y_{t-1} + \tilde{c}(L)\Delta^2\tilde{y}_{t-1}. \end{aligned}$$

Noting that $c_\bullet(0) = \mathcal{A}$ such that $B\mathcal{A}c_\bullet^{-1}(0)B' = I$, it follows that

$$\Gamma = B\mathcal{A} \left\{ c_\bullet^*(1) \begin{pmatrix} -I \\ 0 \\ 0 \end{pmatrix} \beta' + c_\bullet^{-1}(1) \begin{pmatrix} -\beta' + A\beta'_2 \\ -\beta'_1 \\ 0 \end{pmatrix} \right\}$$

and thus

$$\alpha = -B\mathcal{A}c_\bullet^{-1}(1) \begin{pmatrix} I \\ 0 \\ 0 \end{pmatrix}, \quad \alpha'_\perp = - \begin{pmatrix} 0 & I & 0 \\ 0 & 0 & I \end{pmatrix} c_\bullet(1)\mathcal{A}^{-1}B', \quad \alpha'_\perp\Gamma\beta_\perp = \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix}.$$

This confirms the low rank decompositions $\alpha\beta'$ and $\alpha'_\perp\Gamma\beta_\perp = \zeta\eta'$ as well as the rank of these matrices. The VAR(∞) representation then follows from calculating the coefficients of $\tilde{c}(L)$ and $c_\bullet(L)^{-1}$. Therefore, the triangular representation immediately shows that the process $(y_t)_{t \in \mathbb{Z}}$ is I(2) and not integrated of higher order, but leaves the dynamics

contained in $c_{\bullet}(z)$ unspecified. The VAR(∞) representation, on the other hand, focuses on the dynamics, but makes the derivation of the properties of the processes defined by the difference equation less immediate.

To illustrate the function of the constraint (8) to exclude higher order of integration in the solutions process, assume for the moment that the process $(y_t)_{t \in \mathbb{Z}}$ is I(2) and consider $\alpha'_2 \Delta^2 y_t$:

$$\alpha'_2 \Delta^2 y_t = \alpha'_2 \Gamma \Delta y_{t-1} + \alpha'_2 \sum_{j=1}^{\infty} \Pi_j \Delta^2 y_{t-j} + \alpha'_2 \varepsilon_t.$$

Here, $\alpha'_2 \Gamma \beta_{\perp} = 0$ (as α'_2 denotes the second block row of α'_1) such that $\alpha'_2 \Gamma \Delta y_{t-1} = \alpha'_2 \Gamma \beta (\beta' \beta)^{-1} \beta' \Delta y_{t-1}$. Now, using the first difference of the model equation, we see that

$$(\alpha' \alpha)^{-1} \alpha' \Delta^3 y_t = \beta' \Delta y_{t-1} + (\alpha' \alpha)^{-1} (\alpha' \Gamma \Delta^2 y_{t-1} + \alpha' \sum_{j=1}^{\infty} \Pi_j \Delta^3 y_{t-j} + \alpha' \Delta \varepsilon_t)$$

such that $\beta' \Delta y_{t-1} = -(\alpha' \alpha)^{-1} \alpha' \Gamma \Delta^2 y_{t-1} + \Delta v_t$ for some stationary process $(v_t)_{t \in \mathbb{Z}}$.

Combining these facts, we see that

$$\alpha'_2 \Delta^2 y_t = -\alpha'_2 \Gamma \beta (\beta' \beta)^{-1} (\alpha' \alpha)^{-1} \alpha' \Gamma \Delta^2 y_{t-1} + \alpha'_2 \sum_{j=1}^{\infty} \Pi_j \Delta^2 y_{t-j} + \alpha'_2 \varepsilon_t + \gamma \Delta v_t$$

for some matrix γ . Thus (the last equation defines $\tilde{a}(L)$),

$$\alpha'_2 \Delta^2 y_t + \alpha'_2 \Gamma \beta (\beta' \beta)^{-1} (\alpha' \alpha)^{-1} \alpha' \Gamma \Delta^2 y_{t-1} - \alpha'_2 \sum_{j=1}^{\infty} \Pi_j \Delta^2 y_{t-j} = \tilde{a}(L) \Delta^2 y_t = \alpha'_2 \varepsilon_t + \gamma \Delta v_t.$$

Consequently,

$$\tilde{a}(L) \beta_2 \beta'_2 \Delta^2 y_t = \tilde{a}(L) \Delta^2 y_t - \tilde{a}(L) [\beta \beta' + \beta_1 \beta'_1] \Delta^2 y_t = \alpha'_2 \varepsilon_t + \gamma \Delta v_t - \tilde{a}(L) [\beta \beta' + \beta_1 \beta'_1] \Delta^2 y_t.$$

The right hand side process is a stationary process with nonsingular spectrum at $z = 1$ as it equals $\alpha'_2 \varepsilon_t + \Delta \tilde{v}_t$ for stationary process $(\tilde{v}_t)_{t \in \mathbb{Z}}$. This shows that for the properties of $\beta'_2 \Delta^2 y_t$, the square matrix

$$\tilde{a}(1) \beta_2 = \alpha'_2 \left(I_p + \Gamma \beta (\beta' \beta)^{-1} (\alpha' \alpha)^{-1} \alpha' \Gamma - \sum_{j=1}^{\infty} \Pi_j \right) \beta_2$$

is of major importance: if it is nonsingular, then $\beta'_2 \Delta^2 (y_t)_{t \in \mathbb{Z}}$ is stationary. If it is singular such that the process still contains a unit root, then $\beta'_2 \Delta^2 (y_t)_{t \in \mathbb{Z}}$ is an integrated process. Hence, the nonsingularity of (8) ascertains that the solution process $(y_t)_{t \in \mathbb{Z}}$ is I(2) and not I(3) (or integrated of even higher order).

For simplicity of presentation, no deterministic terms are included in the model formulation. Then, the corresponding result for I(2) processes is the main result of this note:

Theorem 1. *Let $(y_t)_{t \in \mathbb{Z}}$ be generated according to Assumption 2. Let the lag length selection be performed by minimizing the information criterion $IC(h; C_T)$ with penalty term $C_T \geq c > 1$, $C_T/T \rightarrow 0$ over the integers $0 \leq h \leq H_T = o(T^{1/3})$. Let the minimal minimizing argument be denoted as $\hat{h}(C_T)$. Then,*

- (i) $\mathbb{P}(\hat{h}(C_T) \leq M) \rightarrow 0$ for each constant $0 < M$.
- (ii) If Assumption 1 holds for Σ_h with function $\theta(h)$, then $\hat{h}(C_T)/h_T^* \rightarrow 1$ in probability, where h_T^* minimizes the function $L_T(h; C_T) = hp^2(C_T - 1)/T + \theta(h)$.
- (iii) If $(y_t)_{t \in \mathbb{Z}}$ is an I(2) invertible VARMA process corresponding to the left-coprime pair $(A(z), B(z))$, then $-2\hat{h}(C_T) \log \rho_0 / \log T \rightarrow 1$ in probability, where $\rho_0^{-1} = \min\{|z| : z \in \mathbb{C}, |B(z)| = 0\}$.

The same results hold if the process is demeaned and/or detrended before estimation.

The result shows that the lag length selection essentially follows the minima of the function $L_T(h; C_T)$ analogously to the stationary case. In the VARMA situation, we obtain a lag length that is proportional to $\log T$ and depends on the location of the zero closest to the unit circle. Otherwise, the minima of $L_T(h; C_T)$ will increase for increasing sample size as the penalty term $hp^2(C_T - 1)/T$ tends to zero for $T \rightarrow \infty$ while $\theta(h)$ decreases monotonously.

The assumptions here are stronger than needed in almost all aspects. Convergence of $\Pi(z)$ on $|z| < 1 + \delta$ can be weakened to $\sum_{j=1}^{\infty} j^a \|\Pi_j\| < \infty$ for suitable $a > 0$. In addition, the iid assumption for $(\varepsilon_t)_{t \in \mathbb{Z}}$ can be weakened to martingale difference type assumptions as in HD. I here use the stronger assumptions in order to stay in line with [Li and Bauer \(2020\)](#) (in the following LB) on which the proof (to be found in Appendix A) is based.⁷

Note that Assumption 2 excludes the finite lag length case in which $(y_t)_{t \in \mathbb{Z}}$ is generated by an autoregression of, say, order h_0 . In this case, we have $\Sigma_h = \Sigma$, $h \geq h_0$. It follows that, asymptotically, $\hat{h}(C_T) \geq h_0$ with probability one if $C_T/T \rightarrow 0$, $C_T \geq c > 1$. Furthermore, the probability of $\hat{h}(C_T) > h_0$ tends to zero for $C_T \rightarrow \infty$ as a function of T as the penalty then dominates the estimation error for $\hat{\Sigma}_h$. This demonstrates that BIC is (weakly) consistent for VAR(h_0) processes.

Inspecting the proof, it is obvious that analogous results also hold for the I(1) and the multifrequency I(1) case. Moreover, often, not only is the lag length selected uniformly for all component processes, but different lag lengths are selected for each component. In addition, in this case, the proof in the appendix can be adapted to show that the asymptotic properties in the integrated case are analogous to the ones in the stationary case obtained by suitably differencing the process.

Such an extension also pertains to VARX models. In this situation, it is also possible to be more general such that the exogenous process does not need to possess a VAR(∞) representation with sufficiently fast decreasing coefficients, but may show (stationary or integrated) long memory if the number of lags to be included is subject to an upper bound. The bottom line in all such cases is that the integration properties of the process are of less concern, if differencing leads to stationarity with sufficiently fast decaying impulse responses.

Finally, note that similar results appear possible for processes of higher order of integration.

3. Simulations

In this section, the theory is illustrated with a simulation exercise involving MA(1) processes of the form:

$$u_t = \varepsilon_t - \theta \varepsilon_{t-1}.$$

for standard Gaussian white noise process $(\varepsilon_t)_{t \in \mathbb{Z}}$. These processes are stationary and invertible for θ being a stable matrix. In the scalar case, the smallest zero of the process equals $\rho_0 = |\theta|$. Therefore, the required number of lags in an autoregressive approximation is controlled via θ and is expected to grow similar to $\log T / (-2 \log \rho_0)$ as a function of the sample size.

A total of $M = 1000$ realizations of the process with sample size $T = 100, 200, 400$, and $T = 800$ are simulated, and one parameter $\theta = -0.9, -0.7, \dots, 0.9$ is varied on a regular grid.

We thus investigate the increase of the average selected lag length for univariate MA(1) processes as a function of the sample size as well as the location of the zero closest to the unit circle. This is performed for the stationary process as well as for a doubly integrated MA(1) process obtained via double summation of the stationary process. The double integration is expected to add two lags compared to the lag length required for the stationary process $(u_t)_{t \in \mathbb{Z}}$ as $\Delta^2 y_t = u_t$. Furthermore, the double integration potentially also adds a deterministic quadratic trend. Thus, lag length selection is performed on the corresponding detrended process in the I(2) case.

Figure 1 shows the resulting average lag lengths selected using BIC. The main effects contained in the theorem are clearly visible: the selected lag length increases with sample size. The logarithmic scale on the x-axis for plot (a) shows⁸ the roughly linear increase in $\log T$. In addition, larger absolute values of θ result in larger lag lengths. The average selected lag lengths with BIC are very similar to the optimal values h_T^* (corresponding to

the stationary process $(u_t)_{t \in \mathbb{Z}}$. The doubly integrated processes $(y_t)_{t \in \mathbb{Z}}$ require roughly two more lags in all cases compared to the stationary process $\Delta^2(y_t)_{t \in \mathbb{Z}}$ except close to $\theta = 0.9$ (which is close to a pole zero cancellation), where only one additional lag results in the lag length selection on average.

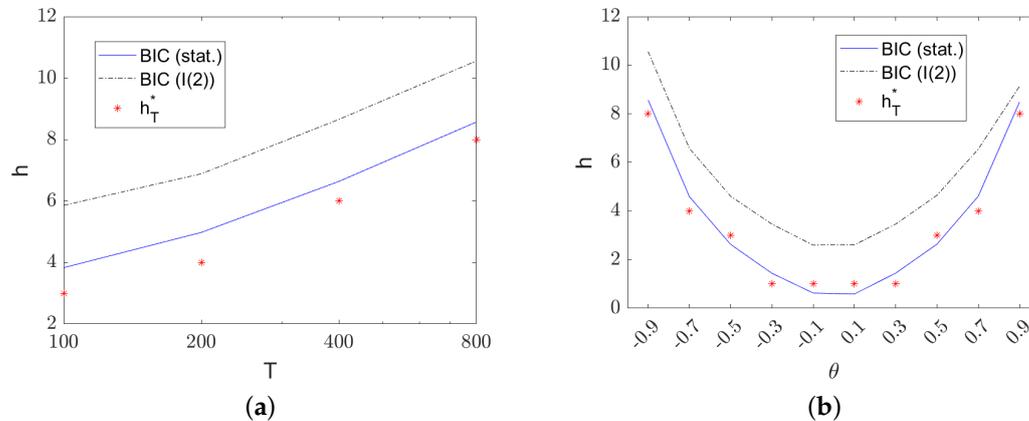


Figure 1. Average selected lag length for the stationary processes (blue) and the $I(2)$ processes (black, dashed). Red stars denote the optimizers h_T^* . (a) For $\theta = -0.9$ over sample sizes $T = 100, 200, 400, 800$. (b) For sample size $T = 800$ over different values of θ .

4. Conclusions

In this paper, we investigate the properties of the commonly used lag length selection using information criteria for VAR approximations of $I(2)$ processes. The discussion shows that the asymptotic properties of the lag length selection criteria are analogous to the properties according to the standard stationary setting, but typically require two extra lags to account for the double integration.

In the invertible VARMA case, this implies that the lag lengths selected using AIC or BIC tend to infinity as a function of the sample size proportional to $\log T$. The proportionality constant depends on the location of the zero closest to the unit circle. This is identical to the stationary case. The proof of the result indicates that this property is robust for a great number of unit roots being present in the data-generating process.

Funding: This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation—Projektnummer 276051388) which is gratefully acknowledged. I acknowledge support for the publication costs by the Deutsche Forschungsgemeinschaft and the Open Access Publication Fund of Bielefeld University.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The author declares no conflict of interest.

Appendix A. Proof of Theorem 1

Proof. As in Johansen (1995), sct. 4.3, it follows that the process

$$u_t = \begin{pmatrix} \tilde{y}_{1,t} - A\Delta\tilde{y}_{3,t} \\ \Delta\tilde{y}_{2,t} \\ \Delta^2\tilde{y}_{3,t} \end{pmatrix} = c_{\bullet}(L)\varepsilon_t, \quad \tilde{y}_t = \begin{pmatrix} \tilde{y}_{1,t} \\ \tilde{y}_{2,t} \\ \tilde{y}_{3,t} \end{pmatrix} = \begin{pmatrix} \beta' \\ \beta'_1 \\ \beta'_2 \end{pmatrix} y_t,$$

is stationary for $A = \bar{\alpha}'\Gamma\beta_2$ and $\beta_1 = \beta_{\perp}\eta$ for appropriate choice of initial conditions y_0, y_1 , and y_2 . The MA representation, as in Johansen (1995), Theorem 4.6, shows that the difference for different values of y_0, y_1 , and y_2 amounts to the addition of deterministic terms, which are easily dealt with.

The properties of $c_{\bullet}(z)$ follow from the usual arguments, as in Johansen (1995), sct. 4.3 (see Theorem 4.6 on p. 58) noting that derivatives of absolutely convergent power series have the same radius of convergence as the original function.

The rest of the proof uses such a triangular representation. Then

$$\begin{aligned} \tilde{y}_{1,t} &= A\Delta\tilde{y}_{3,t} + u_{1,t}, \\ \Delta\tilde{y}_{2,t} &= u_{2,t}, \\ \Delta^2\tilde{y}_{3,t} &= u_{3,t}, \end{aligned}$$

where $\tilde{c}_{\bullet}(L)^{-1}u_t = \varepsilon_t$ is a VAR(∞) process. Note, however, that the triangular representation requires knowledge of β, β_1, β_2 and hence is not operational in general. We use it here only as a technical device in the proof in order to separate processes of different order of integration. In particular, the representation above in conjunction with the nonsingularity of the spectrum of $(u_t)_{t \in \mathbb{Z}}$ at $z = 1$ directly implies that $\tilde{y}_{3,t}$ is I(2) and not cointegrated, $\tilde{y}_{2,t}$ is I(1) and not cointegrated, and $\tilde{y}_{1,t}$ is I(1) or I(0) and cointegrated with $\Delta\tilde{y}_{3,t}$.

Using this notation (A.2) of LB states that⁹

$$\begin{aligned} \Delta^2 y_t = & \Phi_2 \tilde{y}_{2,t-1} + \Phi_3 \tilde{y}_{3,t-1} + \Theta \Delta \tilde{y}_{3,t-1} + \sum_{j=1}^h \Xi_j u_{t-j} + \Xi_{h+1,1:2} \begin{pmatrix} u_{1,t-h-1} \\ u_{2,t-h-1} \end{pmatrix} \\ & + \Xi_{h+2,1} (u_{1,t-h-2} - Au_{3,t-h-1}) + e_t(h+2). \end{aligned}$$

The regressors on the right hand side are a linear transformation of the regressors in the VAR($h + 2$) approximation in levels. Φ_2, Φ_3 , and Θ subsume all nonstationary directions of the regressors. Since these are not cointegrated—as follows from the triangular representation—and all other quantities in the equation are stationary, their coefficients need to be zero, that is, $\Phi_2 = 0, \Phi_3 = 0, \Theta = 0$.

We rewrite the equation as $\Delta^2 y_t = \Psi_u z_t + \Xi_h U_{t-1,h} + e_t(h+2)$, where

$$z_t = (\tilde{y}'_{2,t-1}, \Delta \tilde{y}'_{3,t-1}, \tilde{y}'_{3,t-1})'$$

and $U_{t-1,h}$ collects all stationary regressors. Denoting $\langle a_t, b_t \rangle = T^{-1} \sum_{t=h+1}^T a_t b_t'$ and $W_t = [z_t', U'_{t-1,h}]'$, we obtain, from the Frisch–Waugh–Lovell theorem,

$$\begin{aligned} \hat{\Sigma}_{h+2} &= \langle \Delta^2 y_t, \Delta^2 y_t \rangle - \langle \Delta^2 y_t, W_t \rangle \langle W_t, W_t \rangle^{-1} \langle W_t, \Delta^2 y_t \rangle \\ &= \langle \Delta^2 y_t, \Delta^2 y_t \rangle - \langle \Delta^2 y_t, U_{t-1,h} \rangle \langle U_{t-1,h}, U_{t-1,h} \rangle^{-1} \langle U_{t-1,h}, \Delta^2 y_t \rangle \\ &\quad - \langle \Delta^2 y_t, z_t^\pi \rangle \langle z_t^\pi, z_t^\pi \rangle^{-1} \langle z_t^\pi, \Delta^2 y_t \rangle \end{aligned}$$

where $z_t^\pi = z_t - \langle z_t, U_{t-1,h} \rangle \langle U_{t-1,h}, U_{t-1,h} \rangle^{-1} U_{t-1,h}$. We introduce the scaling matrix $D_T = \text{diag}(T^{-1/2} I_{r+s}, T^{-3/2} I_s)$. Then, $D_T \langle z_t, z_t \rangle D_T \xrightarrow{d} W$ for some random, almost surely positive, definite matrix $W \in \mathbb{R}^{(r+2s) \times (r+2s)}$. Moreover,

$$\|D_T \langle z_t, U_{t-1,h} \rangle \langle U_{t-1,h}, U_{t-1,h} \rangle^{-1} \langle U_{t-1,h}, z_t \rangle D_T\| \leq \|D_T \langle z_t, U_{t-1,h} \rangle\|^2 \|\langle U_{t-1,h}, U_{t-1,h} \rangle\|^{-1}.$$

The theory for stationary processes implies that the second factor is $O_P(1)$. The expectation of the first factor is $O(h/T) = o(1)$ (compare LB, p. 17). Thus, the whole term is $O_P(h/T) = o_p(1)$ uniformly in $h \leq H_T$, implying that $(D_T \langle z_t^\pi, z_t^\pi \rangle D_T)^{-1} = (D_T \langle z_t, z_t \rangle D_T)^{-1} (1 + o_p(1)) = O_P(1)$.

Similarly,

$$\langle \Delta^2 y_t, U_{t-1,h} \rangle \langle U_{t-1,h}, U_{t-1,h} \rangle^{-1} \langle U_{t-1,h}, z_t \rangle D_T = O_P(T^{-1/2})$$

since $\|\langle \Delta^2 y_t, U_{t-1,h} \rangle \langle U_{t-1,h}, U_{t-1,h} \rangle^{-1} - \Psi_h\| = O_P(\sqrt{h/T})$, according to Theorem 7.4.5. of HD, where $\langle \Psi_h U_{t-1,h}, z_t \rangle D_T = O_P(T^{-1/2})$. Since $\langle \Delta^2 y_t, z_t \rangle D_T = O_P(T^{-1/2})$, we obtain $\langle \Delta^2 y_t, z_t^\pi \rangle D_T = O_P(T^{-1/2})$. Thus,

$$\begin{aligned} \langle \Delta^2 y_t, z_t^\pi \rangle \langle z_t^\pi, z_t^\pi \rangle^{-1} \langle z_t^\pi, \Delta^2 y_t \rangle &= (\langle \Delta^2 y_t, z_t^\pi \rangle D_T) (D_T \langle z_t^\pi, z_t^\pi \rangle D_T)^{-1} D_T \langle z_t^\pi, \Delta^2 y_t \rangle \\ &= O_P(T^{-1/2}) O_P(1) O_P(T^{-1/2}) = O_P(1/T). \end{aligned}$$

Comparing with the expression for $\hat{\Sigma}_{h+2}$ from above, we obtain

$$\hat{\Sigma}_{h+2} = \langle \Delta^2 y_t, \Delta^2 y_t \rangle - \langle \Delta^2 y_t, U_{t-1,h} \rangle \langle U_{t-1,h}, U_{t-1,h} \rangle^{-1} \langle U_{t-1,h}, \Delta^2 y_t \rangle + O_P(1/T).$$

Now

$$\Delta^2 \tilde{y}_t = \begin{pmatrix} \Delta^2 u_{1,t} + A \Delta u_{3,t} \\ \Delta u_{2,t} \\ u_{3,t} \end{pmatrix} = \begin{pmatrix} I_r & 0 & A \\ 0 & I_s & 0 \\ 0 & 0 & I_{p-r-s} \end{pmatrix} u_t + \begin{pmatrix} -2u_{1,t-1} + u_{1,t-2} - Au_{3,t-1} \\ -u_{2,t-1} \\ 0 \end{pmatrix}$$

implies

$$|\hat{\Sigma}_{h+2}| = |\hat{\Sigma}_h^u| + O_P(1/T), \quad \hat{\Sigma}_h^u := \langle u_t, u_t \rangle - \langle u_t, U_{t-1,h} \rangle \langle U_{t-1,h}, U_{t-1,h} \rangle^{-1} \langle U_{t-1,h}, u_t \rangle.$$

Thus, $|\hat{\Sigma}_h^u| \geq |\hat{\Sigma}_h| \geq |\hat{\Sigma}_{h+2}^u|$ and $|\hat{\Sigma}_{h+2}| - |\hat{\Sigma}_h^u| = O_P(T^{-1})$ uniformly in $0 < h < H_T$, where $\hat{\Sigma}_h^u$ denotes the estimated residual variance of a long VAR approximation of $(u_t)_{t \in \mathbb{Z}}$ using lag length h .

This is sufficient for the results, as can be seen as follows: divergence to infinity for $\hat{h}(C_T)$ is obvious since the error term is at most of the same order as the penalty term.

To establish (ii), consider the case $C_T \rightarrow \infty$ first. Note that the order $O_P(T^{-1})$ together with $C_T \rightarrow \infty$ implies that $IC(h; C_T) = IC_u(h; C_T)(1 + o_P(1))$, where IC_u denotes the criterion for $(u_t)_{t \in \mathbb{Z}}$. This is the prerequisite for (ii) and (iii) according to the arguments in HD on pp. 332–34 following Theorem 7.4.7. (ii): Let h_T^* denote the minimizer over h of $L_T(h)$ (for $h \in \mathbb{R}_+$). Then, a mean value expansion around h_T^* leads to

$$L_T(\hat{h}_T; C_T) - L_T(h_T^*; C_T) = \frac{1}{2} \theta''(\bar{h}_T) (\hat{h}_T - h_T^*)^2 = \frac{\theta''(\bar{h}_T) (h_T^*)^2}{2\theta(h_T^*)} \left(\frac{\hat{h}_T}{h_T^*} - 1 \right)^2 \theta(h_T^*) \geq 0$$

where \bar{h}_T is an intermediate value. Now $IC(h; C_T)$ is minimized at \hat{h}_T , where $IC(h; C_T) - \hat{\Sigma}_T - L_T(h; C_T) = O_P(T^{-1}) + o_P(\theta(h))$ uniformly in h , according to the above in combination with (7.4.22) of HD. This implies that either $\hat{h}_T/h_T^* \rightarrow 1$ or $\theta(h_T^*) = O_P(T^{-1})$ such that the positive right hand side of the above can be reversed by the estimation error to obtain $IC(\hat{h}_T; C_T) \leq IC(h_T^*; C_T)$.

However, $\theta(h_T^*) = O_P(T^{-1})$ is a contradiction to h_T^* , minimizing $L_T(h; C_T) = hp^2(C_T - 1)/T + \theta(h)$: in that case, $h_T^* - 1$ reduces L_T by $p^2(C_T - 1)/T$ plus a smaller error term due to $C_T \rightarrow \infty$. This leads to $\hat{h}_T/h_T^* \rightarrow 1$.

For $h \rightarrow \infty$, the order $IC(h; C_T) = IC_u(h; C_T)(1 + o_P(1))$ also holds with C_T remaining finite, as in AIC. Since the lag length estimated for finite C_T is at least as large as the one selected using BIC, $\hat{h}_T \rightarrow \infty$ follows, again implying the results.

All statements above also hold for demeaned and/or detrended processes. \square

Notes

- 1 Other variants exist, for example, using the same time points in the summation for all models; compare with Kilian and Lütkepohl (2017).
- 2 Hannan and Deistler (1988) will, in the following, be abbreviated as HD.
- 3 In the unlikely case of draws, the smallest integer h achieving the minimum is selected.
- 4 Lütkepohl and Saikkonen (1999) also correct an argument contained in Ng and Perron (1995).

- 5 Here and below we use the notation X_{\perp} for a full column rank matrix whose columns span the orthonormal complement of the column space of a full column rank matrix X .
- 6 Since we fix the values y_0, y_1 , and y_2 , the processes will only be stationary for appropriate choices of the values, otherwise it is the sum of a stationary process and the effects of the initial values.
- 7 As pointed out by a reviewer (for which I am grateful), the moment condition in HD is slightly stronger than assuming finite fourth moment as used in LB.
- 8 Note that the optima h_T^* correspond to $\log T / (-2 \log \rho_0)$ rounded to the nearest integer.
- 9 LB uses the lag length $h + 2$ instead of h .

References

- Bauer, Dietmar, and Martin Wagner. 2004. *Autoregressive Approximations to MFI(1) Processes*. Working Paper No. 174, Reihe Ökonomie/Economics Series. Vienna: Institut für Höhere Studien (IHS).
- Granger, Clive W. J., and Tae-Hwy Lee. 1989. Investigation of Production, Sales and Inventory relationships using multicointegration and non-symmetric error correction models. *Journal of Applied Econometrics* 4: 145–59. [[CrossRef](#)]
- Hannan, Edward James, and Manfred Deistler. 1988. *The Statistical Theory of Linear Systems*. New York: John Wiley.
- Johansen, Søren. 1995. *Likelihood-Based Inference in Cointegrated Vector Auto-Regressive Models*. Oxford: Oxford University Press.
- Juselius, Katarina. 2006. *The Cointegrated VAR Model: Methodology and Applications*. Oxford: Oxford University Press.
- Kilian, Lutz, and Helmut Lütkepohl. 2017. *Structural Vector Autoregressive Analysis*. Cambridge: Cambridge University Press.
- Li, Yuanyuan, and Dietmar Bauer. 2020. Modeling I (2) processes using vector autoregressions where the lag length increases with the sample size. *Econometrics* 8: 38. [[CrossRef](#)]
- Lütkepohl, Helmut, and Pentti Saikkonen. 1999. *Order Selection in Testing for the Cointegrating Rank of a VAR Process, Cointegration, Causality, and Forecasting*. A Festschrift in Honour of Clive WJ Granger. Oxford: Oxford University Press, pp. 168–99.
- Ng, Serena, and Pierre Perron. 1995. Unit Root Tests in ARMA Models with Data-Dependent Methodes for the Selection of the Truncation Lag. *Journal of the American Statistical Association* 90: 268–81. [[CrossRef](#)]
- Paulsen, Jostein. 1984. Order determination of multivariate autoregressive time series with unit roots. *Journal of Time Series Analysis* 5: 115–27. [[CrossRef](#)]
- Saikkonen, Pentti, and Helmut Lütkepohl. 1996. Infinite-Order Cointegrated Vector Autoregressive Processes: Estimation and Inference. *Econometric Theory* 12: 814–44. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.