# Automatic Inference for Infinite Order Vector Autoregressions

Guido M. Kuersteiner*

April 2004

**Abstract**

Infinite order vector autoregressive (VAR) models have been used in a number of applications ranging from spectral density estimation, impulse response analysis, tests for cointegration and unit roots, to forecasting. For estimation of such models it is necessary to approximate the infinite order lag structure by finite order VAR's. In practice, the order of approximation is often selected by information criteria or by general-to-specific specification tests. Unlike in the finite order VAR case these selection rules are not consistent in the usual sense and the asymptotic properties of parameter estimates of the infinite order VAR do not follow as easily as in the finite order case. In this paper it is shown that the parameter estimates of the infinite order VAR are asymptotically normal with zero mean when the model is approximated by a finite order VAR with a datadependent lag length. The requirement for the result to hold is that the selected lag length satisfies certain rate conditions with probability tending to one. Two examples of selection rules satisfying these requirements are discussed. Uniform rates of convergence for the parameters of the infinite order VAR are also established.

*Massachusetts Institute of Technology, Dept. of Economics, 50 Memorial Drive E52-371A, Cambridge, MA 02142, USA. Email: gkuerste@mit.edu. Web: http://web.mit.edu/gkuerste/www/. Very helpful comments by the editor and two referees led to a substantial improvement of the manuscript. I am particularly indebted to one of the referees for pointing out an error in the proofs. All remaining errors are my own. Financial support from NSF grant SES-0095132 is gratefully acknowledged.

## 1. Introduction

Infinite order vector autoregressive (VAR($\infty$)) models are appealing nonparametric specifications for the covariance structure of stationary processes because they can be justified under relatively weak restrictions on the Wold representation of a stationary process. In practice, the VAR($\infty$) specification needs to be approximated, usually by a VAR(h) model where the truncation parameter $h$ increases with sample size $n$. This approach was proposed by Akaike (1969) and Parzen (1974) for the estimation of spectral densities.

Approximations to VAR($\infty$) models have received renewed interest in recent years in a number of econometric applications. Lütkepohl and Saikkonen (1997) consider impulse response functions in infinite order cointegrated systems. Cointegration tests and inference in systems with infinite order dynamics are considered by Saikkonen and Luukkonen (1997) and Saikkonen and Lütkepohl (1996). Ng and Perron (1995, 2000) use flexible autoregressive specifications in augmented Dickey Fuller (ADF) unit root tests to improve size properties of these tests. Lütkepohl and Poskitt (1996) construct tests for causality using infinite order vector autoregressive processes. Paparoditis (1996), Inoue and Kilian (2002) and Goncalves and Kilian (2004) propose bootstrap procedures for VAR($\infty$) models. Finally, den Haan and Levin (2000) use prewhitening procedures and VAR($\infty$) approximations to estimate heteroskedasticity and autocovariance consistent (HAC) covariance matrices for robust inference. They use AIC and BIC information criteria to select the order of the approximating VAR and report evidence that applying standard kernel based smoothing to estimate spectral densities from the prewhitened residuals does not lead to improvements over estimates that are entirely based on the VAR specification.

The lag length $h$ is the key design parameter in implementing procedures that approximate VAR($\infty$) models. The results of Berk (1974) and Lewis and Reinsel (1985) establish rates of convergence necessary for consistency and asymptotic normality. A number of papers using VAR($h$) approximations do not go beyond listing these restrictions on rates as conditions for their results. In practice, such restrictions however can not be used to construct automated procedures because the lowerbound for the expansion rate of $h$ depends on unknown properties of the data. Moreover, conditions on the growth rate of $h$ as a function of the sample size $n$ are not sufficient to choose $h$ in a finite sample. What is called for are datadependent rules where $h$ is chosen based on information in the sample.

Hannan and Kavalieris (1976) and Hannan and Deistler (1988) analyze the stochastic properties of feasible rules $\hat{h}_n$ based on the AIC and BIC criterion. The AIC information criterion has been shown to posses minimal mean squared error properties for the estimation of parameters in AR($\infty$)

models and minimal integrated mean squared error properties for the estimation of approximations to the spectral density of AR($\infty$) models by Shibata (1980, 1981). Ng and Perron (1995) point out that the AIC criterion violates the conditions on $h$ obtained by Berk (1974) and Lewis and Reinsel (1985). This leads to expansion rates for $h$ that are too slow to eliminate biases that result in shifts of the asymptotic limit distribution of the parameters.

Infinite dimensional models have a long tradition in econometric theory. The work of Sargan (1975) is an early example. The problem of biases caused by parameter spaces that grow in dimension with the sample size has recently been discussed in econometrics by Bekker (1994). Similar effects can be found in various contexts, for example in the work of Donald and Newey (2001), Hahn and Kuersteiner (2002, 2003) and Kuersteiner (2002).

Especially in time series applications finite sample biases can be substantial and may have a dominating effect on inference. Kilian (1998) shows that bootstrap confidence intervals for impulse response functions are severely affected by finite sample biases in the estimates of the underlying autoregressions. He proposes a bias correction to overcome severe distortions in coverage rates. In panel models with lagged dependent variables Hahn, Hausman and Kuersteiner (2000) and Hahn and Kuersteiner (2002) document the predominant effect of finite sample biases on the mean squared error of parameter estimates.

Ng and Perron (1995) propose a general-to-specific testing approach to select the approximate lag order in ADF tests where the underlying model is a VAR($\infty$). Their work extends results of Hall (1994) for lag order selection in ADF tests when the underlying model is a finite order VAR to the infinite order case. Ng and Perron (1995) advocate general-to-specific selection rules to overcome the problems of AIC and BIC in selecting the lag-length in VAR($\infty$) approximations described above although their focus is on the performance of unit root tests and not on the estimation of the VAR($\infty$) parameters. They show that distributional properties of ADF tests are not affected by biases induced by AIC and BIC but report simulation evidence demonstrating the advantages in terms of finite sample size of the ADF tests when parametrized with their lag selection procedure.

In this paper the results of Ng and Perron (1995) are extended to estimation and inference in VAR($\infty$) models. It is argued that the convergence properties of $\hat{h}_n$ based on model selection procedures typically are not strong enough to apply the arguments of Eastwood and Gallant (1991) for admissible estimation. This is true for the general-to-specific approach of Ng and Perron (1995) as well as for conventional model selection methods based on information criteria. In fact, in infinite dimensional parameter spaces adaptiveness of selection rules, a concept that has appeared in the literature and will be defined more precisely in Section 2, is hard to show. Moreover, the results of Shibata (1980,1981) do not establish the asymptotic distribution of parameter estimates in an AR($\hat{h}_n$) model when $\hat{h}_n$ is selected by AIC. Such a result seems to be missing in the literature to this date.

Here, the arguments do not rely on adaptiveness properties of the selection rule. An alternative proof, based on the work of Lewis and Reinsel (1985) is used to show that $h$ can be replaced by $\hat{h}_n$ determined by the general-to-specific approach of Ng and Perron (1995) without affecting the limiting distribution of the parameters in the VAR($h$) approximation. This leads to fully automated approximations to the VAR($\infty$) model that do not suffer from higher order biases as approximations using AIC and BIC generally would. Nevertheless, in the special case where the underlying process is a VARMA model, a modification of AIC also can be used without affecting the limiting distribution, a result that is discussed at the end of Section 2. Uniform rates of convergence for the parameters of the VAR($\hat{h}_n$) approximation are also obtained. These rates in turn can be used to establish rates for functionals of the VAR parameters such as the spectral density matrix.

The main results of the paper are presented in Section 2, Section 3 contains some conclusions and all the proofs are collected in Section 4.

## 2. Linear Time Series Models

Let $y_t \in \mathbb{R}^p$ be a strictly stationary time series with an infinite order moving average representation

$$y_t = \mu_y + C(L)v_t. \tag{2.1}$$

Here, $\mu_y \in \mathbb{R}^p$ is a constant and $v_t$ is a strictly stationary and conditionally homoskedastic martingale difference sequence. The lag polynomial $C(L)$ is defined as $C(L) = \sum_{j=0}^{\infty} C_j L^j$ where $L$ is the lag operator.

**Assumption A.** *Let $v_t \in \mathbb{R}^p$ be strictly stationary and ergodic, with $E(v_t|\mathcal{F}_{t-1}) = 0$, $E(v_t v_t'|\mathcal{F}_{t-1}) = \Sigma_v$ where $\Sigma_v$ is a positive definite symmetric matrix of constants. Let $v_t^i$ be the i-th element of $v_t$, $\xi = (\xi_1, ..., \xi_k) \in \mathbb{R}^k$ and $v = (v_{t_1}^{i_1}, ..., v_{t_k}^{i_k})$ such that $\phi_{i_1,...i_k,t_1,...,t_k}(\xi) = Ee^{i\xi' v}$ is the joint characteristic function with corresponding joint k-th order cumulant function defined as $\operatorname{cum}^*_{i_1,...,i_k}(t_1, ..., t_k) = \frac{\partial^{u_1+...+u_k}}{\partial \xi_1^{u_1} ... \partial \xi_k^{u_k}}|_{\xi=0} \ln \phi_{i_1,...i_k,t_1,...,t_k}(\xi)$ where $u_i$ are nonnegative integers such that $u_1 + ... + u_k = k$. By stationarity it is enough to define $\operatorname{cum}_{i_1,...,i_k}(t_1, ..., t_{k-1}) = \operatorname{cum}^*_{i_1,...,i_k}(t_1, ..., t_{k-1}, 0)$. Assume that*

$$\sum_{t_1=-\infty}^{\infty} ... \sum_{t_{k-1}=-\infty}^{\infty} |\operatorname{cum}_{i_1,...,i_k}(t_1, ..., t_{k-1})| < \infty \tag{2.2}$$

*where the sum converges for all $k \leq 4$ and all $i_j \in \{1, ..., p\}$ with $j \in \{1, ..., k\}$.*

Assumption A is weaker than the assumptions imposed in Lewis and Reinsel (1985) where independence of the innovations is assumed but is somewhat stronger than the assumptions in Hannan and Deistler (1988, Theorem 7.4.8) who also allow for the more general heteroskedastic case which is excluded here by the requirement that $E(v_t v_t' | \mathcal{F}_{t-1}) = \Sigma_v$. Recently, Goncalves and Kilian (2003) have obtained explicit formulas for the norming constant when the innovations are conditionally heteroskedastic. The summability Assumption (2.2) is quite common in the literature on HAC-estimation. Andrews (1991) for example uses a similar condition and shows that (2.2) is implied by a mixing condition.

**Assumption B.** *The lag polynomial $C(L)$ with coefficient matrices $C_j$ satisfies $C_0 = I_p$, $\sum_{j=1}^{\infty} j^{1/2} \|C_j\| < \infty$ where $\|A\|^2 = \operatorname{tr} AA'$ for a square matrix $A$ and $\det C(z) \neq 0$ for $|z| \leq 1$ where $z \in \mathbb{C}$.*

**Assumption C.** *For $C_j$ as defined in Assumption (B) it holds that $\sum_{j=1}^{\infty} j^2 \|C_j\| < \infty$ and $\det C(z) \neq 0$ for $|z| \leq 1$.*

The summability restriction on the impulse coefficients $C_j$ in Assumptions B and C is stronger than the condition imposed by Lewis and Reinsel (1985) where only $\sum_{j=0}^{\infty} \|C_j\| < \infty$ is required. It is needed here to achieve similar flexibility in the central limit theorem as Lewis and Reinsel (1985). Assumption B implies that $y_t$ has an infinite order VAR representation given by

$$y_t = \mu + \sum_{j=1}^{\infty} \pi_j y_{t-j} + v_t \tag{2.3}$$

where $\mu = C(1)^{-1}\mu_y$ and $C(L)^{-1} = \pi(L)$ with $\pi(L) = I - \sum_{j=1}^{\infty} \pi_j L^j$. The impulse response function $C(L)$ of $y_t$ is thus a functional of $\pi(L)$ defined by $C(L) = \pi(L)^{-1}$. Another functional of interest is the spectral density $f_y(\lambda)$ of $y_t$ where $f_y(\lambda) = (2\pi)^{-1} \pi(e^{i\lambda})^{-1}\Sigma_v \left(\pi(e^{-i\lambda})^{-1}\right)'$. For inferential purposes we are often interested in $f_y(0)$, the spectral density at frequency zero.

The VAR($\infty$) representation in (2.3) needs to be approximated in practice by a model with a finite number of parameters, in the case considered here a VAR(h) model. The approximate model with VAR coefficient matrices $\pi_{1,h}, ...\pi_{h,h}$ is thus given by

$$y_t = \mu_{y,h} + \pi_{1,h} y_{t-1} + ... + \pi_{h,h} y_{t-h} + v_{t,h} \tag{2.4}$$

where $\Sigma_{v,h} = E v_{t,h} v_{t,h}'$ is the mean squared prediction error of the approximating model.

It was shown by Berk (1974) and Lewis and Reinsel (1985) that the parameters $(\pi_{1,h},...,\pi_{h,h})$ are root-n consistent and asymptotically normal for $\pi(h) = (\pi_1,...,\pi_h)$ in an appropriate sense to be made explicit later if $h$ does not increase too quickly, i.e. if $h$ is chosen such that $h^3/n \to 0$. At the same time $h$ must not increase too slowly to avoid asymptotic biases. Berk (1974) shows that $h$ needs to increase such that $n^{1/2}\sum_{j=h+1}^{\infty}\pi_j \to 0$ as $h, n \to \infty$. In practice such rules are difficult to implement as they only determine rates of expansion for $h$ and do not lead directly to feasible selection criteria for $h$. Ng and Perron (1995) argue that information criteria such as the Akaike criterion do not satisfy[1] the conditions of Berk (1974) and Lewis and Reinsel (1985). In general these criteria can not be used to choose $h$ if asymptotic unbiasedness, as measured by the location of the asymptotic limiting distribution, is desired. More specifically, if $h$ is such that $n^{1/2}\sum_{j=h+1}^{\infty}\pi_j \to c \neq 0$ then bias terms due to asymptotic misspecification of the model are of order $n^{-1/2}$. These biases are more severe than the usual finite sample biases that are typically of order $n^{-1}$.

To avoid the problems that arise from using information criteria to select the order of the approximating model we use the sequential testing procedure analyzed in Ng and Perron (1995). Let $\pi(h) = (\pi_1',...,\pi_h')'$, $Y_{t,h} = \left(y_t' - \bar{y}',...,y_{t-h+1}' - \bar{y}'\right)'$ where $\bar{y} = n^{-1}\sum_{t=1}^{n} y_t$ and $M_h = \sum_{t=h+1}^{n} Y_{t-1,h}Y_{t-1,h}'$. Define $M_h^{-1}(1)$ to be the lower-right $p \times p$ block of $M_h^{-1}$. Let $\Gamma_h$ be the $hp \times hp$ matrix whose $(m,n)$th block is $\Gamma_{n-m}^{yy}$ and $\Gamma_{1,h}' = \left[\Gamma_{-1}^{yy},...,\Gamma_{-h}^{yy}\right]$ where $\Gamma_{j-i}^{yy} = \mathrm{Cov}(y_{t-i}, y_{t-j}')$. The coefficients of the approximate model satisfy the equations $(\pi_{1,h},...,\pi_{h,h}) = \Gamma_{1,h}\Gamma_h^{-1}$. Let $\hat{\Gamma}_{1,h} = (n-h)^{-1}\sum_{t=h}^{n-1} Y_{t,h}\left(y_{t+1} - \bar{y}\right)'$ and $\hat{\Gamma}_h = (n-h)^{-1}\sum_{t=h}^{n-1} Y_{t,h}Y_{t,h}'$. The estimated error covariance matrix is $\hat{\Sigma}_{v,h} = (n-h)^{-1}\sum_{t=h+1}^{n}\hat{v}_{t,h}\hat{v}_{t,h}'$ where $\hat{v}_{t,h} = y_t - \hat{\pi}_{1,h}y_{t-1} - ... - \hat{\pi}_{h,h}y_{t-h}$ with coefficients

$$\hat{\pi}(h)' = \hat{\Gamma}_{1,h}'\hat{\Gamma}_h^{-1}.$$

Under Assumptions A and B it follows from Hannan and Deistler (1988, Theorem 7.4.6) that $\hat{\Sigma}_{v,h} \to \Sigma_{v,h}$ uniformly in $h \leq h_{\max}$ and $h_{\max} = o((n/\log n)^{1/2})$. A Wald test for the null hypothesis that the coefficients of the last lag $h$ are jointly 0 is then, in Ng and Perron's notation,

$$J(h,h) = (\mathrm{vec}\,\hat{\pi}_{h,h})'\left(\hat{\Sigma}_{v,h} \otimes M_h^{-1}(1)\right)^{-1}(\mathrm{vec}\,\hat{\pi}_{h,h}).$$

The following lag order selection procedure from Ng and Perron (1995) is adopted.

**Definition 2.1.** *The general-to-specific procedure chooses i)* $\hat{h}_n = h$ *if, at significance level* $\alpha$, $J(h,h)$ *is the first statistic in the sequence* $J(i,i), \{i = h_{\max},...,1\}$, *which is significantly different from zero or*

[1] A special case where a version of AIC satisfies Berk's conditions is discussed at the end of this Section.

*ii)* $\hat{h}_n = 0$ *if* $J(i,i)$ *is not significantly different from zero for all* $i = h_{\max}, ..., 1$ *where* $h_{\max}$ *is such that* $h_{\max}^3/n \to 0$ *and* $n^{1/2} \sum_{j=h_{\max}+1}^{\infty} \|\pi_j\| \to 0$ *as* $n \to \infty$.

Implementation of the general-to-specific procedure may be difficult in practice because the critical values depend on complicated conditional densities which are not Gaussian in the parameters and therefore not $\chi^2$ for the test statistics. This seems to be the case even though the underlying joint and marginal densities can be assumed to be Gaussian with easily estimated coefficients[2]. For a discussion of these issues see Sen (1979), and in particular Pötscher (1991) and Leeb and Pötscher (2003). Note that Lemma 3 of Pötscher (1991) does not hold in the present context. This means that the sequence of test statistics $J(i,i), \{i = h_{\max}, ..., 1\}$ is not asymptotically independent and thus the conditional density of $J(i,i)$ is not the same as the marginal density. Whether numerical methods or the bootstrap could be used to obtain an operational version of the general-to-specific approach is beyond the scope of this paper.

In order to illustrate the problems with establishing results that allow to substitute $h$ with $\hat{h}_n$ in $\hat{\pi}(h)$ we consider the lag order estimate $\hat{h}_n$ based on the AIC and BIC criteria. The lag order estimate is defined as $\hat{h}_n = \arg\min \hat{Q}_n(h)$ with $\hat{Q}_n(h) = \log\det\hat{\Sigma}_{v,h} + hp^2 C_n/n$ where $C_n = 2$ for AIC and $C_n = \log n$ for BIC. Hannan and Deistler (1988, Theorem 7.4.7) show under slightly different assumptions than here, that $\hat{Q}_n(h)$ can be essentially replaced by $Q_n(h) = hp^2/n(C_n - 1) + \operatorname{tr}\left(\Sigma^{-1}\left(\Sigma_{v,h} - \Sigma\right)\right)$. Shibata (1980)[3] argues that $Q_n(h)$ can be interpreted as the one step ahead squared prediction error obtained from predicting $y_t$ with an AR(h) model. Misspecification bias manifests itself in the term $\Sigma_{v,h} - \Sigma$ that depends amongst other things on the dimension $p$ of $y_t$ and affects the choice of $h$.[4] Define $h_n^* = \arg\min Q_n(h)$ as the optimal lag order minimizing the squared prediction error. In the context of VARMA models which are special cases of (2.3) the results of Hannan and Kavalieris (1984, 1986) imply that if $\hat{h}_n$ is selected by AIC or BIC then $\hat{h}_n - h_n^* = o_p(\sqrt{\log n})$. Eastwood and Gallant (1991) and Ng and Perron (1995) define the concept of adaptive selection rules. A sequence of random variables $\hat{a}_n$ is an adaptive selection rule if there is a deterministic rule $a_n$ such that $\hat{a}_n - a_n = o_p(1)$. The discussion of AIC and BIC based selection rules $\hat{h}_n$ shows that these rules are not adaptive for $h_n^*$ in the sense of Eastwood and Gallant (1991) and Ng and Perron (1995).

[2] I am grateful to one of the referees for pointing out this fact.

[3] Hannan and Deistler (1988, p.317) discuss this interpretation.

[4] Abadir, Hadri and Tzavalis (1999) analyze non-stationary VAR's where the asymptotic limiting distribution of OLS estimators is also shifted away from the origin. They find an explicit relation between the dimension $p$ and the bias. The situation there is however quite different from the one considered here where bias is due to misspecification.

Similarly, the results of Ng and Perron (1995) imply that $\hat{h}_n$ selected by the procedure in Definition (2.1) satisfies $P\left(h_{\min} \leq \hat{h}_n \leq h_{\max}\right) \to 1$ as $n \to \infty$ for any sequence $h_{\min}$ such that $h_{\min} \leq h_{\max}$ and $h_{\max} - h_{\min} \to \infty$. Such a result again is not strong enough to guarantee that $\hat{h}_n$ is adaptive for $Mh_{\max}$ where $M$ is an arbitrary positive constant. Any argument that relies on the adaptiveness property of selection rules to establish that $\hat{\pi}(\hat{h}_n)$ has the same asymptotic properties as $\hat{\pi}(Mh_{\max})$ therefore can not be applied. It may be possible to prove adaptivness properties of selection rules but such results do not seem to be readily available in the literature.

For this reason an alternative proof strategy is chosen here. The following weaker consequence of Lemma 5.2 of Ng and Perron (1995) which follows directly from their proof turns out to be sufficient to establish the feasibility of a fully automatic approximation to the VAR($\infty$) model.

**Lemma 2.2.** *Let $\hat{h}_n$ be given by Definition (2.1). Let $h_{\min}$ be any sequence such that $h_{\max} \geq h_{\min}$, $h_{\max} - h_{\min} \to \infty$ and $n^{1/2} \sum_{j=h_{\min}+1}^{\infty} \|\pi_j\| \to 0$ as $n \to 0$. Then $\lim_n P(\hat{h}_n \leq h_{\min}) = 0$.*

The following two main results of this paper establish that the results in Lewis and Reinsel (1985) essentially remain valid if in $\hat{\pi}(h)$, $h$ is replaced by $\hat{h}_n$. The proofs establish uniform convergence of $\hat{\pi}(h)$ over a set $H_n$ of values $h$ such that $\hat{h}_n$ is contained in $H_n$ with probability tending to one. First, an asymptotic normality result is established for an arbitrary but absolutely summable linear transformation $l(h)$ of the parameters into the real line. In particular this result implies that arbitrary finite linear combinations of elements in $\hat{\pi}(\hat{h}_n)$ are asymptotically normal. By the Cramér-Wold theorem this also implies that any finite combination of elements in $\hat{\pi}(\hat{h}_n)$ is jointly asymptotically normal.

**Theorem 2.3.** *Let $\hat{h}_n$ be given by Definition (2.1). i) Let Assumptions (A) and (B) hold. Let $l(h) = (l_1', ..., l_h')'$ be the $p^2h \times 1$ section of an infinite dimensional vector $l$ such that for some constants $M_1$ and $M_2$, $0 < M_1 \leq \sum_{j=1}^{\infty} \|l_j\| \leq \sqrt{M_2} < \infty$. Let $\omega_h = l(h)' \left(\Gamma_h^{-1} \otimes \Sigma_v\right) l(h)$. Then $\lim_{h\to\infty} \omega_h = \omega$ exists and is bounded and*

$$\sqrt{n} l(\hat{h}_n)' \left[\mathrm{vec}(\hat{\pi}(\hat{h}_n)' - \pi(\hat{h}_n)')\right] \xrightarrow{d} N\left(0, \omega\right).$$

*ii) Instead of Assumption (B) let Assumption (C) hold. Let $h_{\max}$ be as in Definition (2.1), $h_{\min}$ be defined as in Lemma (2.2) with $\Delta_n \equiv h_{\max} - h_{\min} \to \infty$, $\Delta_n = O(n^{\delta})$ for $0 < \delta < 1/3$ and assume that there exists some $\underline{h}$ such that $\underline{h} \leq h_{\min}$, $\Delta_n/(h_{\min} - \underline{h}) \to 0$ and $\underline{h} \to \infty$, some $h^{**}$ such that $h^{**} \leq h_{\min}$, $h^{**} \to \infty$ and $\Delta_n/\sqrt{h^{**}} \to 0$ and a sequence $l(h) = (l_{1,h}', ..., l_{h,h}')'$ of $p^2h \times 1$ vectors partitioned into $p^2 \times 1$ vectors $l_{j,h}$ such that for some constants $M_1$ and $M_2$, $0 < M_1 \leq \|l(h)\|^2 = l(h)'l(h) \leq M_2 < \infty$*

*for all* $h = 1, 2, ...,$ *and* $\sum_{j=1}^{h^{**}} \|l_{j,h} - l_{j,h_{\max}}\|^2 = o\left(\Delta_n^{-2}\right)$, $\sum_{j=1}^{h} j^2 \|l_{j,h} - l_{j,h_{\max}}\|^2 < \infty$ *as well as* $\sum_{j=\underline{h}+1}^{h} \|l_{j,h}\|^2 = o(\Delta_n^{-2})$ *for all* $h \to \infty$, $h_{\min} \leq h \leq h_{\max}$. *Then*

$$\sqrt{n}\frac{l(\hat{h}_n)'}{\omega_{\hat{h}_n}} \left[\mathrm{vec}(\hat{\pi}(\hat{h}_n)' - \pi(\hat{h}_n)')\right] \xrightarrow{d} N(0,1).$$

**Remark 1.** *The rate at which* $\Delta_n \to \infty$ *can essentially be arbitrarily slow. Thus the restrictions on* $h^{**}$ *and* $\underline{h}$ *are quite weak.*

**Remark 2.** *Note that the tail summability conditions in the second part of the theorem are automatically satisfied for the fixed vectors* $l$ *with* $l'l < \infty$ *that satisfy the additional constraint* $\sum_{j=\underline{h}+1}^{\infty} \|l_j\|^2 = o\left(\Delta_n^{-2}\right)$ *for some* $\underline{h} \to \infty$. *The second part allows for more general limit theorems where* $l(h)$ *fluctuates except in the 'tails'.*

**Remark 3.** *While the theorem essentially provides the same results as Lewis and Reinsel (1985) for many cases of practical interest it nevertheless requires somewhat stronger assumptions both on* $C_i$ *and* $l(h)$. *A different proof strategy may lead to different and maybe less restrictive conditions but it seems unlikely that a result at the same level of generality as in Lewis and Reinsel (1985) can be shown without establishing adaptiveness of* $\hat{h}_n$.

**Remark 4.** *For i) it also follows that* $\sqrt{n}\left(l(\hat{h}_n)'\,\mathrm{vec}(\hat{\pi}(\hat{h}_n)') - l'\,\mathrm{vec}\,(\pi(\infty)')\right) \xrightarrow{d} N(0,\omega)$ *because* $|l'\,\mathrm{vec}\,(\pi(\infty)')| \leq \left|l(\hat{h}_n)'\,\mathrm{vec}(\pi(\hat{h}_n)')\right| + \sqrt{M_2}\sum_{j=h_{\min}+1}^{\infty} \|\pi_j\|$ *with* $\sqrt{n}\sum_{j=h_{\min}+1}^{\infty} \|\pi_j\| \to 0.$

The next result is a refined version of Theorem 1 of Lewis and Reinsel (1985). It establishes a uniform rate of convergence for the parameter estimates when the lag length is chosen by the general-to-specific approach of Ng and Perron (1995).

**Theorem 2.4.** *Let Assumptions (A) and (B) hold. Let* $\hat{h}_n$ *be given by Definition (2.1). Then*

$$\left\|\hat{\pi}(\hat{h}_n) - \pi(\hat{h}_n)\right\| = O_p((\log n/n)^{1/2})$$

*and* $\sum_{j=\hat{h}_n+1}^{\infty} \|\pi_j\| = o_p\left((\log n/n)^{1/2}\right)$.

The result in Theorem (2.4) is particularly useful to establish consistency and convergence rates of functionals of $\pi(L)$ such as the spectral density matrix of $y_t$. The result presented here is stronger than a corresponding result for nonstochastic lag order selection presented in Lewis and Reinsel (1985, Theorem 1) where only uniform consistency is established without specifying the convergence rate.

Theorem (2.4) complements results in Hannan and Deistler (1988, Theorem 7.4.5) where the case of nonstochastic $h$ sequences is analyzed.

Theorems (2.3) and (2.4) do not rely on a specific model selection procedure. All that is required for the theorems to apply is that there are sequences $h_{\min}$ and $h_{\max}$ satisfying the conditions stated previously and a datadependent rule $\hat{h}_n$ such that $P\left(h_{\min} \leq \hat{h}_n \leq h_{\max}\right) \to 1$ as $n \to \infty$. It is thus quite plausible that feasibility can be established for a broader class of selection procedures than the one considered here.

Under more restrictive assumptions this can even be done for AIC based procedures. In fact for VARMA models $\sum_{j=h+1}^{\infty} \pi_j = O(\rho_0^{-h})$ where $\rho_0$ is the modulus of a zero of $C(z)$ nearest $|z| = 1$. Hannan and Deistler (1988, Theorem 6.6.4 and p.334) show that $\hat{h}_n$ selected by AIC satisfies $\hat{h}_n/h_n^* - 1 = o_p(1)$ for $h_n^* = \log n/\left(2\log\rho_0\right)$. It thus follows that $n^{1/2}\sum_{j=Mh_n^*+1}^{\infty} \pi_j = O(n^{1/2-M/2})$ which is $o(1)$ for $M > 1$. This suggests that at least for VARMA systems AIC could be used as an automatic order selection criterion for autoregressive approximations[5]. Feasibility of this approach follows from Theorems (2.3) and (2.4) because there exist $h_{\min} = Mh_n^*/2$ and $h_{\max} = (\log n)^a$, $1 < a < \infty$, satisfying the requirements of the theorems. This shows that for $M > 2$ and $\hat{h}_n$ selected by AIC, the rule $M\hat{h}_n$ can be used instead of the general-to-specific procedure if the underlying model is a VARMA model.

## 3. Conclusions

In this paper data-dependent selection rules for the specification of VAR($h$) approximations to VAR($\infty$) models are analyzed. It is shown that the method of Ng and Perron (1995) can be used to produce a datadependent selection rule $\hat{h}_n$, such that the parameters of the approximating VAR($h$) model are asymptotically normal for the parameters of the underlying VAR($\infty$) model. The asymptotic normality result does only hold on essentially finite subsets of the parameter space. Uniform rates of convergence for the VAR($\infty$) parameters are thus obtained in addition.

The results presented here extend the existing literature where so far model selection has been carried out mostly in terms of information criteria. Such criteria are known to result in sizeable higher order biases. On the other hand, the selection criteria analyzed here do not suffer from these biases. The paper also reconsiders some existing proof strategies in the context of infinite dimensional parameter spaces where the concept of consistent model selection is hard to apply.

[5] I am grateful to one of the referees for pointing out this fact which is discussed in Hannan and Deistler (1988, p. 262).

## 4. Proofs

### 4.1. Auxiliary Lemmas

The following Lemmas are used in the proof of Theorem 2.3. The matrix norm $\|A\|_2^2 = \sup_{l \neq 0} l'A'Al/l'l$, known as the two-norm, is adopted from Lewis and Reinsel (1985, p.396) where the less common notation $\|.\|_1$ is used. There it is also shown that for two matrices $A$ and $B$, the inequalities $\|AB\|^2 \leq \|A\|_2^2 \|B\|^2$ and $\|AB\|^2 \leq \|A\|^2 \|B\|_2^2$ hold. First it is shown that the mean of $y_t$ can be replaced by an estimate without affecting the asymptotics.

**Lemma 4.1.** *Let* $\check{Y}_{t,h} = \left(y_t' - \mu_y', y_{t-1}' - \mu_y', ..., y_{t-h+1}' - \mu_y'\right)'$ *and let* $\check{\Gamma}_{1,h} = (n-h)^{-1} \sum_{t=h}^{n-1} \check{Y}_{t,h} \left(y_{t+1} - \mu_y\right)'$ *and* $\check{\Gamma}_h = (n-h)^{-1} \sum_{t=h}^{n-1} \check{Y}_{t,h} \check{Y}_{t,h}'$. *Let* $\check{\pi}(\hat{h}_n)' = \check{\Gamma}_{1,\hat{h}_n}' \check{\Gamma}_{\hat{h}_n}^{-1}$ *and* $\hat{\pi}(\hat{h}_n)$ *is defined in Section (2). Then* $\left\|\hat{\pi}(\hat{h}_n) - \check{\pi}(\hat{h}_n)\right\| = o_p(n^{-1/2})$.

**Proof.** Choose $\delta$ such that $0 < \delta < 1/3$ and pick a sequence $h_{\min}^*$ such that $h_{\max} \geq h_{\min}^*$, $h_{\max} - h_{\min}^* \to \infty$, $h_{\max} - h_{\min}^* = O(n^\delta)$ and $n^{1/2} \sum_{j=h_{\min}^*+1}^{\infty} \|\pi_j\| \to 0$. Define

$$h_{\min} = \max\left\{h_{\min}^*, h_{\max} - \left(n^{1/2} \sum_{j=h_{\min}^*+1}^{\infty} \|\pi_j\|\right)^{-1}\right\}.$$

It follows that $h_{\min} \leq h_{\max}$ and

$$\min\left\{h_{\max} - h_{\min}^*, \left(n^{1/2} \sum_{j=h_{\min}^*+1}^{\infty} \|\pi_j\|\right)^{-1}\right\} = h_{\max} - h_{\min} \leq \left(n^{1/2} \sum_{j=h_{\min}^*+1}^{\infty} \|\pi_j\|\right)^{-1} \tag{4.1}$$

such that $h_{\max} - h_{\min} \to \infty$. Because $h_{\max} - h_{\min}^* = O\left(n^\delta\right)$ it follows that $h_{\max} - h_{\min} = O(n^\delta)$. Since $h_{\min} \in [h_{\min}^*, h_{\max}]$ it also follows that $n^{1/2} \sum_{j=h_{\min}+1}^{\infty} \|\pi_j\| \to 0$ and $\left(h_{\min}\right)^3/n \to 0$. Let $H_n = \{h | h_{\min} \leq h \leq h_{\max}\}$. Note that from Lemma (2.2) it follows that $\hat{h}_n \in H_n$ with probability tending to one. Consider

$$\left\|\hat{\pi}(\hat{h}_n) - \check{\pi}(\hat{h}_n)\right\| \leq \left\|\hat{\Gamma}_{1,\hat{h}_n} - \check{\Gamma}_{1,\hat{h}_n}\right\| \left\|\hat{\Gamma}_{\hat{h}_n}^{-1}\right\|_2 + \left\|\check{\Gamma}_{1,\hat{h}_n}\right\|_2 \left\|\hat{\Gamma}_{\hat{h}_n}^{-1} - \check{\Gamma}_{\hat{h}_n}^{-1}\right\|$$

where

$$\begin{aligned}\hat{\Gamma}_{1,\hat{h}_n} - \check{\Gamma}_{1,\hat{h}_n} &= \left(n - \hat{h}_n\right)^{-1} \sum_{t=\hat{h}_n}^{n-1} \check{Y}_{t,\hat{h}_n} \left(\mu_y - \bar{y}\right)' + \left(\mathbf{1}_{\hat{h}_n} \otimes \left(\mu_y - \bar{y}\right)\right) \left(n - \hat{h}_n\right)^{-1} \sum_{t=\hat{h}_n}^{n-1} \left(y_{t+1} - \mu_y\right)' \\ &\quad + \frac{n - 1 - \hat{h}_n}{n - \hat{h}_n} \left(\mathbf{1}_{\hat{h}_n} \otimes \left(\mu_y - \bar{y}\right)\right) \left(\mu_y - \bar{y}\right)'\end{aligned}$$

It now can be established that

$$
\begin{aligned}
E \max_{h \in H_n} \left\| (n-h)^{-1} \sum_{t=h}^{n-1} \left(y_{t+1} - \mu_y\right) \right\|^2 &\leq \sum_{h=h_{\min}}^{h_{\max}} (n-h)^{-2} \sum_{t,s=h}^{n-1} \operatorname{tr} E \left(y_{t+1} - \mu_y\right) \left(y_{s+1} - \mu_y\right)' \\
&\leq O(n^{\delta}) \left(n - h_{\max}\right)^{-2} \sum_{t,s=h_{\min}}^{n-1} \left| \operatorname{tr} E \left(y_{t+1} - \mu_y\right) \left(y_{s+1} - \mu_y\right)' \right| \\
&= O\left(n^{-1+\delta}\right).
\end{aligned}
$$

and

$$
\begin{aligned}
E \max_{h \in H_n} \left\| (n-h)^{-1} \sum_{t=h}^{n-1} \check{Y}_{t,h} \right\|^2 &= \sum_{h=h_{\min}}^{h_{\max}} (n-h)^{-2} \operatorname{tr} \sum_{t,s=h}^{n-1} E \check{Y}_{t,h} \check{Y}'_{s,h} \\
&\leq \Delta_n \frac{h_{\max}}{\left(n - h_{\max}\right)^2} \sum_{t,s=h_{\min}}^{n-1} \operatorname{tr} \left\| \Gamma_{t-s}^{yy} \right\| \\
&= o\left(n^{-2/3+\delta}\right).
\end{aligned}
$$

Furthermore, $\left\| \mathbf{1}_{\hat{h}} \otimes (\mu_y - \bar{y}) \right\|^2 = o_p\left(n^{-2/3}\right)$ such that $\left\| \hat{\Gamma}_{1,\hat{h}} - \check{\Gamma}_{1,\hat{h}} \right\| = o_p(n^{-1/2})$ by the Markov inequality and Lemma (2.2). In the same way,

$$
\begin{aligned}
\hat{\Gamma}_{\hat{h}_n} - \check{\Gamma}_{\hat{h}_n} &= \left(n - \hat{h}_n\right)^{-1} \sum_{t=\hat{h}_n}^{n-1} \check{Y}_{t,\hat{h}_n} \left(\mathbf{1}_{\hat{h}_n} \otimes (\mu_y - \bar{y})\right)' + \left(\mathbf{1}_{\hat{h}} \otimes (\mu_y - \bar{y})\right) \left(n - \hat{h}_n\right)^{-1} \sum_{t=\hat{h}_n}^{n-1} \check{Y}'_{t,\hat{h}_n} \\
&+ \frac{n - 1 - \hat{h}_n}{n - \hat{h}_n} \left(\mathbf{1}_{\hat{h}_n} \otimes (\mu_y - \bar{y})\right) \left(\mathbf{1}_{\hat{h}_n} \otimes (\mu_y - \bar{y})\right)'
\end{aligned}
$$

such that $\left\| \hat{\Gamma}_{\hat{h}_n} - \check{\Gamma}_{\hat{h}_n} \right\| = o_p(n^{-2/3+\delta})$. By the arguments in the proof of Theorem 1 in Lewis and Reinsel (1985) it then also follows that $\left\| \hat{\Gamma}_{\hat{h}_n}^{-1} - \check{\Gamma}_{\hat{h}_n}^{-1} \right\| = o_p(n^{-1/2})$. ∎

**Lemma 4.2.** *If $y_t$ has typical element $a$ denoted by $y_t^a$ and $w_{t,i} = \left(y_t - \mu_y\right) \left(y_{t-i} - \mu_y\right)'$ then*

$$
E(w_{t,i} w'_{s,j}) = \Gamma_i^{yy} \Gamma_{-j}^{yy} + \gamma_{t-i+j-s}^{yy} \Gamma_{t-s}^{yy} + \Gamma_{t-s+j}^{yy} \Gamma_{t-i-s}^{yy} + \mathcal{K}_4
$$

*where the scalar coefficient $\gamma_{t-s}^{yy}$ is defined as $\gamma_{t-s}^{yy} = E\left(y_t - \mu_y\right)' \left(y_s - \mu_y\right)$ and $\mathcal{K}_4$ is a $p \times p$ matrix with typical element $(a, b)$ denoted by $[.]_{a,b}$ and given as*

$$
\left[\mathcal{K}_4(t, s, i, j)\right]_{a,b} = \sum_{r=1}^{p} \operatorname{cum}_{a,b,r,r}^{y*}(s, t, t-i, s-j).
$$

*where $\operatorname{cum}_{a,b,r,r}^{y*}(s, t, t-i, s-j)$ is defined as*

$$
\begin{aligned}
&\operatorname{cum}_{a,b,r,r}^{y*}(s, t, t-i, s-j) = \operatorname{cum}_{a,b,r,r}^{y*}(j, t-s+j, t-i+j-s, 0) \\
= &\sum_{j_1=1,..,j_4=1}^{p} \sum_{l_1=0,...,l_4=0}^{\infty} c_{l_1}^{a,j_1} c_{l_2}^{b,j_2} c_{l_3}^{r,j_3} c_{l_4}^{r,j_4} \operatorname{cum}_{j_1,...,j_4}(l_4 - l_1 + j, l_4 - l_2 + t - s + j, l_4 - l_3 + t - i + j - s)
\end{aligned}
$$

*with* $c_j^{a,b} = [C_j]_{a,b}$*. It also follows that*

$$\sum_{s,t,j=-\infty}^{\infty} \left| \mathrm{cum}^{y*}_{a,b,r,r}(s,t,t-i,s-j) \right| < \infty.$$

**Proof.** Without loss of generality assume $\mu_y = 0$. Then the matrix $w_{t,i}w'_{s,j} = y_t y'_s y'_{t-i} y_{s-j} = y_t y'_s \sum_{r=1}^{p} y^r_{t-i} y^r_{s-j}$ has typical element $(a,b)$ equal to $y^a_t y^b_s \sum_{r=1}^{p} y^r_{t-i} y^r_{s-j}$. The result follows from applying $E(wxyz) = E(wx)E(yz) + E(wy)E(xz) + E(wz)E(xy) + \mathrm{cum}^*(x,y,w,z)$ for any set of scalar random variables $x,y,w,z$ with $E(x) = 0$ and $E|x|^4 < \infty$ with the same conditions on $y, w$ and $z$. It thus follows that

$$\begin{aligned} E\left[w_{t,i}w'_{s,j}\right]_{a,b} &= \sum_{r=1}^{p} E\left(y^a_t y^b_s y^r_{t-i} y^r_{s-j}\right) \\ &= \gamma^{yy}_{t-i+j-s}\left[\Gamma^{yy}_{t-s}\right]_{a,b} + \left[\Gamma^{yy}_{i}\Gamma^{yy}_{-j}\right]_{a,b} + \left[\Gamma^{yy}_{t-s+j}\Gamma^{yy}_{t-i-s}\right]_{a,b} + \sum_{r=1}^{p} \mathrm{cum}^{y*}_{a,b,r,r}(s,t,t-i,s-j). \end{aligned}$$

For the summability of the cumulant note that

$$\sum_{t_1=-\infty}^{\infty} \cdots \sum_{t_3=-\infty}^{\infty} \left|\mathrm{cum}_{i_1,\dots,i_k}(l_1+t_1-t_4,\dots,l_4+t_3-t_4)\right| < \infty$$

uniformly in $l_1, \dots l_4$ by Assumption (A). The result then follows from the absolute summability of $c_j^{a,b}$ for $a,b \in \{1,\dots,p\}$. ■

**Corollary 4.3.** *Let* $K_{pp}$ *be the* $p^2 \times p^2$ *commutation matrix* $K_{pp} = \sum_{i,j=1}^{p} e_i e'_j \otimes e_j e'_i$ *where* $e_i$ *is the* $i$*-th unit* $p$*-vector. If* $y_t$ *is a vector of* $p$ *random variables then*

$$E(y_t \otimes y_s)\left(y'_r \otimes y'_q\right) = \Gamma^{yy}_{t-r} \otimes \Gamma^{yy}_{s-q} + \mathrm{vec}\,\Gamma^{yy}_{s-t}\left(\mathrm{vec}\,\Gamma^{yy}_{q-r}\right)' + \left(\Gamma^{yy}_{t-q} \otimes \Gamma^{yy}_{s-r}\right)K_{pp} + \mathcal{K}_4(t,s,r,q)$$

*where* $\mathcal{K}_4(t,s,r,q)$ *is a matrix with*

$$[\mathcal{K}_4(t,s,r,q)]_{a,b} = \mathrm{cum}^{y*}_{\lceil a/p\rceil, a \bmod p, \lceil b/p\rceil, b \bmod p}(s,t,r,q)$$

$\lceil a \rceil$ *is the smallest integer larger than* $a$ *and* $a \bmod p = 0$ *is interpreted as* $a \bmod p = p$ *with*

$$\sum_{t_1,\dots,t_3=-\infty}^{\infty} |\mathcal{K}_4(t_1,t_2,t_3,t_4)| < \infty.$$

**Proof.** Note that $(y_t y'_r \otimes y_s y'_q) = \mathrm{vec}\, y_s y'_t (\mathrm{vec}\, y_q y'_r)' = \mathrm{vec}\, y_s y'_t \left(\mathrm{vec}\, y_r y'_q\right)' K_{pp} = (y_t y'_q \otimes y_s y'_r) K_{pp}$. ■

**Lemma 4.4.** *Let $H_n$ be defined as in Lemma (4.1) and let Assumptions (A) hold. Then*

$$E\left(\max_{h\in H_n}\left\|\check{\Gamma}_h-\Gamma_h\right\|^2\right)=O\left(n^{-2/3+\delta}\right).$$

**Proof.** Without loss of generality assume that $\mu_y=0$. Then

$$\begin{aligned}E\left(\max_{h\in H_n}\left\|\check{\Gamma}_h-\Gamma_h\right\|^2\right) &\leq \textstyle\sum_{h=h_{\min}}^{h_{\max}} E\left\|\check{\Gamma}_h-\Gamma_h\right\|^2\\ &\leq \textstyle\sum_{h=h_{\min}}^{h_{\max}} c_0hp^2/(n-h)\leq \Delta_n c_0\dfrac{h_{\max}p^2}{n-h_{\max}}=O\left(n^{-2/3+\delta}\right)\end{aligned}$$

because

$$\begin{aligned}E\left\|\check{\Gamma}_h-\Gamma_h\right\|^2 &= \frac{1}{(n-h)^2}\sum_{t,s=h}^{n-1}\sum_{j_1,j_2=1}^{h}\operatorname{tr}E\left(y_{t-j_1}y_{t-j_2}'-\Gamma_{j_1-j_2}^{yy}\right)\left(y_{s-j_1}y_{s-j_2}'-\Gamma_{j_1-j_2}^{yy}\right)'\\ &= \frac{1}{(n-h)^2}\sum_{t,s=h}^{n-1}\sum_{j_1,j_2=1}^{h}\operatorname{tr}\left(\gamma_{t-s}^{yy}\Gamma_{t-s}^{yy}+\Gamma_{t-j_2+j_1-s}^{yy}\Gamma_{t-j_1-s+j_2}^{yy}+\mathcal{K}_4\right)=O\left(h/(n-h)\right)\end{aligned}$$

by Lemma 4.2 ■

**Lemma 4.5.** *Let Assumption (A) hold and assume that $\sum_{j=1}^{\infty}\|C_j\|<\infty$. Then $\|\Gamma_h\|_2<\infty$ and $\left\|\Gamma_h^{-1}\right\|_2<\infty$ uniformly in $h$. Let $x_h\in\mathbb{R}^h$ with $\limsup_h\|x_h\|<\infty$. Then $\|\Gamma_hx_h\|<\infty$ and $\left\|\Gamma_h^{-1}x_h\right\|<\infty$. Let $\Gamma_h^{j,k}$ be the $j,k$-th block of $\Gamma_h^{-1}$. Then, $\sum_{j=1}^{h}\left\|\Gamma_{k-j}^{yy}\right\|^2<\infty$ and $\sum_{j=1}^{h}\left\|\Gamma_h^{j,k}\right\|^2<\infty$ uniformly in $k,h$. Moreover, $\sup_{j,k}\left\|\Gamma_h^{j,k}\right\|<\infty$ uniformly in $h$.*

**Proof.** The properties $\|\Gamma_h\|_2<\infty$ and $\left\|\Gamma_h^{-1}\right\|_2<\infty$ follow from Berk (1974, p. 493) and Lewis and Reinsel (1985, p.397). Then, $\|\Gamma_hx_h\|\leq\|x_h\|\|\Gamma_h\|_2<\infty$ and $\left\|\Gamma_h^{-1}x_h\right\|\leq\|x_h\|\left\|\Gamma_h^{-1}\right\|_2<\infty$. For the last statement take $e_{k,h}'=(0_p,...,0_p,I_p,0_p,...)'$ where the $p\times p$ identity matrix $I_p$ is at the $k$-th block. It follows that $\|e_{k,h}\|^2=p$ which is uniformly bounded in $h$, $\sum_{j=1}^{h}\left\|\Gamma_{k-j}^{yy}\right\|^2=\|\Gamma_he_{k,h}\|<\sqrt{p}\|\Gamma_h\|_2<\infty$ with a similar argument holding for $\sum_{j=1}^{h}\left\|\Gamma_h^{j,k}\right\|^2$. The last assertion follows from $\left\|\Gamma_h^{j,k}\right\|=\left\|e_{j,h}'\Gamma_h^{-1}e_{k,h}\right\|\leq p\left\|\Gamma_h^{-1}\right\|_2<\infty$ for any $j,k$ uniformly in $h$. ■

**Lemma 4.6.** *Let Assumptions (A) and (C) hold. Then, $\sum_{k=1}^{h}k^2\left\|\Gamma_k^{yy}\right\|<\infty$ and $\sum_{k=1}^{h}\left\|\Gamma_h^{j,k}\right\|<\infty$ uniformly in $j,h$.*

**Proof.** The first statement follows immediately from Assumption (C). The second result follows from Hannan and Deistler (1988, Theorem 6.6.11). ■

**Lemma 4.7.** *Let Assumptions (A) and (C) and the conditions of Theorem 2.3 hold. Then,*

$$\sum_{k=h^*}^{h}\sum_{j=1}^{\underline{h}}\left\|\Gamma_h^{j,k}\right\| = o((h_{\min}-\underline{h})^{-1})$$

*for all $h^*, h$ such that $h \geq h_{\min} \geq h^* \geq \underline{h}$, $h^*-\underline{h}= O(h_{\min}-\underline{h})$, $h_{\min}-h^* = O(h_{\min}-\underline{h})$ and $\underline{h}\to\infty$. If instead of (C), Assumption (B) holds then for any fixed constant $k_0<\infty$ and for $h_{\min}, h_{\max}$ as defined in Lemma 4.1 it follows that $\sup_{h\in[h_{\min},h_{\max}]}\sup_{j\leq k_0}\sum_{k=h_{\min}+1}^{h}\left\|\Gamma_h^{j,k}\right\|\to 0$ as $n\to\infty$ where the sum is assumed to be zero for all $n$ where $k_0 \geq h_{\min}$.*

**Proof.** Let $\Gamma_\infty$ be the infinite dimensional matrix with j,k-th block $\Gamma_{k-j}^{yy}$ for $j,k=1,2,....$and $\Gamma_\infty^{-1}$ the inverse of $\Gamma_\infty$ with $j,k$-th block denoted by $\Gamma_\infty^{j,k}$. From Lewis and Reinsel (1985, p.401) and Hannan and Deistler (1988, Theorem 7.4.2) it follows that

$$\begin{aligned}\sum_{k=h^*}^{h}\sum_{j=1}^{\underline{h}}\left\|\Gamma_\infty^{j,k}\right\| &\leq \sum_{k=h^*}^{h}\sum_{j=1}^{\underline{h}}\sum_{i=0}^{j-1}\|\pi_i\|\left\|\Sigma_v^{-1}\right\|\|\pi_{i+k-j}\| \\ &\leq \frac{\left\|\Sigma_v^{-1}\right\|}{h^*-\underline{h}}\left(\sum_{r=h^*-\underline{h}}^{\infty}r^2\|\pi_r\|\right)\left(\sum_{s=1}^{\infty}\|\pi_s\|\right) = o((h^*-\underline{h})^{-1}).\end{aligned} \tag{4.2}$$

where $\pi_j=0$ for $j<0$. Next use the bound $\left\|\Gamma_h^{j,k}\right\|\leq\left\|\Gamma_\infty^{j,k}\right\|+\left\|\Gamma_h^{j,k}-\Gamma_\infty^{j,k}\right\|$. From Lewis and Reinsel (1985, p.402) it follows that

$$\begin{aligned}\sum_{k=h^*}^{h}\sum_{j=1}^{\underline{h}}\left\|\Gamma_h^{j,k}-\Gamma_\infty^{j,k}\right\| &\leq \sum_{k=h^*}^{h}\sum_{j=1}^{\underline{h}}\sum_{i=0}^{j-1}\|\pi_{i,i+h-j}-\pi_i\|\left\|\Sigma_{v,i+h-j}^{-1}\right\|\|\pi_{i+k-j,i+h-j}\| \\ &+\sum_{k=h^*}^{h}\sum_{j=1}^{\underline{h}}\sum_{i=0}^{j-1}\|\pi_i\|\left\|\Sigma_{v,i+h-j}^{-1}\right\|\|\pi_{i+k-j,i+h-j}-\pi_{i+k-j}\| \\ &+\sum_{k=h^*}^{h}\sum_{j=1}^{\underline{h}}\sum_{i=0}^{j-1}\|\pi_i\|\left\|\Sigma_{v,i+h-j}^{-1}-\Sigma_v^{-1}\right\|\|\pi_{i+k-j}\|\end{aligned}$$

where the last term is $o((h^*-\underline{h})^{-1})$ because $\left\|\Sigma_{v,i+h-j}^{-1}-\Sigma_v^{-1}\right\| = O(1)$ and the same argument as in (4.2) applies. Next, note that

$$\begin{aligned}&\sum_{k=h^*}^{h}\sum_{j=1}^{\underline{h}}\sum_{i=0}^{j-1}\|\pi_{i,i+h-j}-\pi_i\|\left\|\Sigma_{v,i+h-j}^{-1}\right\|\|\pi_{i+k-j,i+h-j}\| \\ \leq\; &\sum_{j=1}^{\underline{h}}\sum_{i=0}^{j-1}\|\pi_{i,i+h-j}-\pi_i\|\left\|\Sigma_{v,i+h-j}^{-1}\right\|\sum_{k=h^*}^{h}\|\pi_{i+k-j,i+h-j}-\pi_{i+k-j}\| \\ &+\sum_{k=h^*}^{h}\sum_{j=1}^{\underline{h}}\sum_{i=0}^{j-1}\|\pi_{i,i+h-j}-\pi_i\|\left\|\Sigma_{v,i+h-j}^{-1}\right\|\|\pi_{i+k-j}\|\end{aligned}$$

where the second term again is $o((h^*-\underline{h})^{-1})$ because of the uniform bound in (4.3). From Hannan and Deistler (1988, Theorem 6.6.12 and p. 336) it follows that

$$\begin{aligned}\sum_{k=h^*}^{h}\sum_{j=1}^{\underline{h}}\|\pi_{i+k-j,i+h-j}-\pi_{i+k-j}\| &\leq \sum_{j=1}^{\underline{h}}\sum_{k=h_{\min}+i-j+1}^{\infty}\|\pi_k\| \\ &\leq \sum_{j=1}^{\underline{h}}(h_{\min}-j)^{-2}\sum_{k=h_{\min}-\underline{h}}^{\infty}k^2\|\pi_k\| \\ &= o\left(\sum_{j=h_{\min}-\underline{h}}^{h_{\min}-1}j^{-2}\right) = o((h_{\min}-\underline{h})^{-1})\end{aligned}$$

where the bound holds uniformly in $i = 1, 2, ....$ Similarly, for all $i < j$,

$$\|\pi_{i,i+h-j} - \pi_i\| \leq \sum_{s=1}^{i+h-j} \|\pi_{s,i+h-j} - \pi_s\| \leq \sum_{s=i+h_{\min}-\underline{h}}^{\infty} \|\pi_s\|$$

where the first inequality follows from the fact that $h - j \geq 0$ for all $h \in H_n$ and $j \leq \underline{h}$ and the second inequality follows from $h \geq h_{\min}$ and Hannan and Deistler (1988, Theorem 6.6.12 and p. 336). Substituting for $\|\pi_{i,i+h-j} - \pi_i\|$ it can then be seen that uniformly for $j \leq \underline{h}$,

$$(4.3) \qquad \begin{aligned} \sum_{i=0}^{j-1} \|\pi_{i,i+h-j} - \pi_i\| &\leq \sum_{i=0}^{j-1} \sum_{s=i+h_{\min}-\underline{h}}^{\infty} \|\pi_s\| \\ &\leq \sum_{i=0}^{\infty} (i + h_{\min} - \underline{h})^{-2} \sum_{s=i+h_{\min}-\underline{h}}^{\infty} s^2 \|\pi_s\| \\ &\leq \sum_{i=h_{\min}-\underline{h}}^{\infty} i^{-2} \sum_{s=h_{\min}-\underline{h}}^{\infty} s^2 \|\pi_s\| = o\left((h_{\min} - \underline{h})^{-1}\right). \end{aligned}$$

This shows that $\sum_{k=h^*}^{h} \sum_{j=1}^{\underline{h}} \left\|\Gamma_h^{j,k} - \Gamma_\infty^{j,k}\right\| = o\left((h_{\min} - \underline{h})^{-1}\right)$.

For the second part note that

$$\begin{aligned} \sup_{j \leq k_0} \sum_{k=h_{\min}+1}^{h} \left\|\Gamma_\infty^{j,k}\right\| &\leq \sum_{j=1}^{k_0} \sum_{k=h_{\min}+1}^{h} \left\|\Gamma_\infty^{j,k}\right\| \\ &\leq k_0 \left\|\Sigma_v^{-1}\right\| \left(\sum_{r=h_{\min}-k_0}^{\infty} \|\pi_r\|\right) \left(\sum_{s=1}^{\infty} \|\pi_s\|\right) \to 0 \end{aligned}$$

uniformly in $h$. Also, note that

$$\sum_{i=0}^{j-1} \|\pi_{i,i+h-j} - \pi_i\| \leq \sum_{i=0}^{k_0} \sum_{s=h_{\min}-k_0}^{\infty} \|\pi_s\| \to 0$$

as well as

$$\sum_{j=1}^{k_0} \sum_{k=h_{\min}+1}^{h} \|\pi_{i+k-j,i+h-j} - \pi_{i+k-j}\| \leq \sum_{j=0}^{k_0} \sum_{k=h_{\min}+1-k_0}^{\infty} \|\pi_k\| \to 0$$

such that $\sum_{k=h_{\min}+1}^{h} \sum_{j=1}^{k_0} \left\|\Gamma_h^{j,k} - \Gamma_\infty^{j,k}\right\| \to 0$ uniformly in $h$ by the same arguments as before. ■

**Lemma 4.8.** *Let Assumptions (A) and (C) hold, define*

$$\Gamma_{h_{\max}}^{-1} = \begin{bmatrix} \left[\Gamma_{h_{\max}}^{-1}\right]_{11} & \left[\Gamma_{h_{\max}}^{-1}\right]_{12} \\ \left[\Gamma_{h_{\max}}^{-1}\right]_{21} & \left[\Gamma_{h_{\max}}^{-1}\right]_{22} \end{bmatrix}, \Gamma_{h_{\max}} = \begin{bmatrix} \Gamma_{11,h_{\max}} & \Gamma_{12,h_{\max}} \\ \Gamma_{21,h_{\max}} & \Gamma_{22,h_{\max}} \end{bmatrix}$$

*such that* $\left[\Gamma_{h_{\max}}^{-1}\right]_{11}$ *is the right upper* $hp \times hp$ *block of* $\Gamma_{h_{\max}}^{-1}$ *and similarly for* $\Gamma_{11,h_{\max}}$ *and let* $A = \Gamma_{12,h_{\max}}\Gamma_{22,h_{\max}}^{-1}\Gamma_{21,h_{\max}}$ *with typical* $p \times p$ *block* $(a, b)$ *denoted by* $A_{a,b}$. *Then*

$$\sum_{a=1}^{h^*} \sum_{b=1}^{h} \|A_{a,b}\| = o\left((h_{\min} - h^*)^{-1}\right)$$

*for any sequence* $h^*$ *such that* $h_{\min} - h^* \to \infty$ *as* $h \to \infty$.

**Proof.** Note that $A_{a,b} = \sum_{j_1,j_2=1}^{h_{\max}-h} \Gamma^{yy}_{j_1+h-a} \Gamma^{j_1,j_2}_{h_{\max}-h} \Gamma^{yy}_{b-j_2-h}$ because $\Gamma_{22,h_{\max}} = \Gamma_{11,h_{\max}-h}$ by the Toeplitz structure of the covariance matrix. Then it follows by Lemma (4.6) that

$$\begin{aligned}\sum_{a=1}^{h^*}\sum_{b=1}^{h}\|A_{a,b}\| &\leq \sum_{a=1}^{h^*}\sum_{j_1=1}^{h_{\max}-h}\left\|\Gamma^{yy}_{j_1+h-a}\right\| \sup_{j_1\leq h_{\max}-h}\sum_{j_2=1}^{h_{\max}-h}\left\|\Gamma^{j_1,j_2}_{h_{\max}-h}\right\|\sum_{b=-\infty}^{\infty}\left\|\Gamma^{yy}_{b}\right\| \\ &\leq c\left(\sum_{j=1+h_{\min}-h^*}^{h_{\max}-1} j\left\|\Gamma^{yy}_{j}\right\|\right) = o\left((h_{\min}-h^*)^{-1}\right)\end{aligned}$$

where $\sup_{j_1\leq h}\sum_{j_2=1}^{h}\left\|\Gamma^{j_1,j_2}_{h}\right\|\sum_{b=-\infty}^{\infty}\left\|\Gamma^{yy}_{b}\right\| < c < \infty$ uniformly in $h$. ■

### 4.2. Proof of Main Theorems

**Proof of Theorem 2.3:.** The proof is identical for parts i) and ii) unless otherwise stated. Define $h_{\min}$ and $H_n$ as in the proof of Lemma (4.1). In view of Lemma (4.1) we can assume that $\mu_y = 0$. We therefore set $\bar{y} = 0$. Let $w_{0n}(h) = l(h)'\left[\operatorname{vec}(\hat{\pi}(h)' - \pi(h)')\right]$. Note that $\sqrt{n}w_{0n}(h_{\max})/\omega_{h_{\max}} \xrightarrow{d} N(0,1)$ by Hannan and Deistler (1985, Theorem 7.4.8). Then

$$\sqrt{n}\left(\frac{w_{0n}(\hat{h}_n)}{\omega_{\hat{h}_n}} - \frac{w_{0n}(h_{\max})}{\omega_{h_{\max}}}\right) = \sqrt{n}\left(\frac{\omega_{h_{\max}}\left(w_{0n}(\hat{h}_n) - w_{0n}(h_{\max})\right) + w_{0n}(h_{\max})\left(\omega_{h_{\max}} - \omega_{\hat{h}_n}\right)}{\omega_{\hat{h}_n}\omega_{h_{\max}}}\right)$$

where $\omega_h$ is uniformly bounded from below and above by Lewis and Reinsel (1985, p.400) such that $\omega_{\hat{h}_n}$ is bounded from below and above with probability one. It is thus enough to show that

$$\sqrt{n}\left(w_{0n}(\hat{h}_n) - w_{0n}(h_{\max})\right) = o_p(1) \tag{4.4}$$

and

$$\omega_{h_{\max}} - \omega_{\hat{h}_n} = o_p(1). \tag{4.5}$$

Next note that for any $\eta > 0$,

$$\begin{aligned}P\left[\left|\sqrt{n}\left(w_{0n}(\hat{h}_n) - w_{0n}(h_{\max})\right)\right| > \eta\right] &\leq P\left[\max_{h\in H_n}\left|\sqrt{n}\left(w_{0n}(h) - w_{0n}(h_{\max})\right)\right| > \eta\right] \\ &\quad + P\left[\hat{h}_n \notin H_n\right]\end{aligned}$$

where the second probability goes to zero by Lemma (2.2).

Let $u_{t,h} = y_t - \sum_{j=1}^{h}\pi_j y_{t-j}$. From Lewis and Reinsel (1985, Equation 2.7) it follows that

$$w_{0n}(h) = l(h)'\operatorname{vec}\left((n-h)^{-1}\sum_{t=h}^{n-1} u_{t+1,h}Y'_{t,h}\hat{\Gamma}_h^{-1}\right)$$

such that

$$\begin{aligned}
&\sqrt{n-h_{\max}}\left(w_{0n}(h)-w_{0n}(h_{\max})\right)\\
=\;&\sqrt{n-h_{\max}}\left((n-h)^{-1}-(n-h_{\max})^{-1}\right)l(h)'\operatorname{vec}\left(\textstyle\sum_{t=h}^{n-1}u_{t+1,h}Y_{t,h}'\hat{\Gamma}_h^{-1}\right)\\
&+(n-h_{\max})^{-1/2}\textstyle\sum_{t=h}^{n-1}l(h)'\left(\hat{\Gamma}_h^{-1}Y_{t,h}'\otimes I_p\right)\operatorname{vec}\left((u_{t+1,h}-u_{t+1,h_{\max}})\right)\\
&+(n-h_{\max})^{-1/2}l(h)'\left(\hat{\Gamma}_h^{-1}\otimes I_p\right)\operatorname{vec}\left(\textstyle\sum_{t=h}^{n-1}u_{t+1,h_{\max}}Y_{t,h}'-\sum_{t=h_{\max}}^{n-1}u_{t+1,h_{\max}}Y_{t,h}'\right)\\
&+(n-h_{\max})^{-1/2}\left(\textstyle\sum_{t=h_{\max}}^{n-1}\left(l(h)'\left(\hat{\Gamma}_h^{-1}Y_{t,h}\otimes I_p\right)-l(h_{\max})'\left(\hat{\Gamma}_{h_{\max}}^{-1}Y_{t,h_{\max}}\otimes I_p\right)\right)u_{t+1,h_{\max}}\right)\\
=\;&w_{4n}+w_{5n}+w_{6n}+w_{7n}
\end{aligned}$$

where $w_{4n},...,w_{7n}$ are defined in the obvious way. First, consider

$$\begin{aligned}
|w_{4n}| &\leq \left|\sqrt{n-h_{\max}}\left((n-h)^{-1}-(n-h_{\max})^{-1}\right)\right|\|l(h)\|\left\|\hat{\Gamma}_h^{-1}\right\|_2\textstyle\sum_{t=h}^{n-1}\|u_{t+1,h}\|\,\|Y_{t,h}\|\\
&\leq \frac{\Delta_n}{(n-h_{\max})^{3/2}}\left\|\hat{\Gamma}_h^{-1}\right\|_2\textstyle\sum_{t=h}^{n-1}\|u_{t+1,h}\|\,\|Y_{t,h}\|.
\end{aligned}$$

In order to establish a bound for $\max_{h\in H_n}|w_{4n}|$ we consider $\max_{h\in H_n}\left\|\hat{\Gamma}_h^{-1}\right\|_2$, $\max_{h\in H_n}\|u_{t+1,h}\|$ and $\max_{h\in H_n}\|Y_{t,h}\|$ in turn.

From Lewis and Reinsel (1985,p.397) we have $Z_{h,n}=\left\|\hat{\Gamma}_h^{-1}-\Gamma_h^{-1}\right\|_2/F\left(\left\|\hat{\Gamma}_h^{-1}-\Gamma_h^{-1}\right\|_2+F\right)\leq\left\|\hat{\Gamma}_h-\Gamma_h\right\|_2$ where $F$ is a constant such that $\left\|\Gamma_h^{-1}\right\|_2\leq F$ uniformly in $h$ and $E\left\|\hat{\Gamma}_h-\Gamma_h\right\|_2^2\leq E\left(\max_{h\in H_n}\left\|\hat{\Gamma}_h-\Gamma_h\right\|^2\right)=O\left(n^{-2/3+\delta}\right)$ by Lemma (4.4) such that $\max_{h\in H_n}Z_{h,n}=o_p(n^{-1/3+\delta/2})$. Then, $\max_{h\in H_n}\left\|\hat{\Gamma}_h^{-1}-\Gamma_h^{-1}\right\|_2=\max_{h\in H_n}F^2Z_{h,n}/(1-FZ_{h,n})=o_p(n^{-1/3+\delta/2})$ and

$$\max_{h\in H_n}\left\|\hat{\Gamma}_h^{-1}\right\|_2\leq F+\max_{h\in H_n}\left\|\hat{\Gamma}_h^{-1}-\Gamma_h^{-1}\right\|_2=O_p(1). \tag{4.6}$$

For $\max_{h\in H_n}\|u_{t+1,h}\|$ consider

$$\begin{aligned}
E\left(\max_{h\in H_n}\|u_{t+1,h}\|\right) &= E\left(\max_{h\in H_n}\left\|y_{t+1}-\textstyle\sum_{j=1}^{h}\pi_j y_{t+1-j}\right\|\right)\\
&\leq E\|y_{t+1}\|+E\left(\max_{h\in H_n}\textstyle\sum_{j=1}^{h}\|\pi_j\|\,\|y_{t+1-j}\|\right)\\
&\leq E\|y_{t+1}\|+\textstyle\sum_{j=1}^{\infty}\|\pi_j\|\,E\|y_{t+1-j}\|<\infty
\end{aligned}$$

such that $\max_{h\in H_n}\|u_{t+1,h}\|=O_p(1)$ by the Markov inequality. Finally, $\|Y_{t,h}\|^2=\sum_{j=0}^{h-1}\|y_{t-j}\|^2$ such that

$$E\max_{h\in H_n}\|Y_{t,h}\|^2=h_{\max}E\|y_t\|^2=o(n^{1/3}).$$

These results show that

$$\max_{h\in H_n}|w_{4n}| = o_p\left(\Delta_n n^{1/6}(n-h_{\max})^{-1/2}\right) = o_p\left(n^{-1/3+\delta}\right)$$

For $|w_{5n}|$ consider $w_{5n} = w_{51n} + w_{52n}$ where

$$w_{51n} = (n-h_{\max})^{-1/2}\sum_{t=h}^{n-1} l(h)'\left(\left(\hat{\Gamma}_h^{-1}-\Gamma_h^{-1}\right)\otimes I_p\right)\text{vec}\left(\left(u_{t+1,h}-u_{t+1,h_{\max}}\right)Y_{t,h}'\right)$$

and

$$w_{52n}(h) = (n-h_{\max})^{-1/2}\sum_{t=h}^{n-1} l(h)'\left(\Gamma_h^{-1}Y_{t,h}\otimes I_p\right)\text{vec}\left(u_{t+1,h}-u_{t+1,h_{\max}}\right).$$

For $w_{51n}$ consider

$$|w_{51n}| \leq \max_{h\in H_n}\left\|l(h)'\left(\left(\hat{\Gamma}_h^{-1}-\Gamma_h^{-1}\right)\otimes I_p\right)\right\|(n-h_{\max})^{-1/2}\sum_{t=h_{\min}}^{n-1}\max_{h\in H_n}\left\|\left(u_{t+1,h}-u_{t+1,h_{\max}}\right)Y_{t,h}'\right\|$$

where $E\max_{h\in H_n}\left\|\left(u_{t+1,h}-u_{t+1,h_{\max}}\right)Y_{t,h}'\right\| \leq \sum_{j=h_{\min}+1}^{h_{\max}}\|\pi_j\|\left(E\|y_{t-j}\|^2\right)^{1/2}\left(E\max_{h\in H_n}\|Y_{t,h}\|^2\right)^{1/2}$ with $\sup_t E\|y_t\|^2 \leq c < \infty$ and

$$E\max_{h\in H_n}\|Y_{t,h}\|^2 = o(n^{1/3})$$

from before such that

$$\begin{aligned}(n-h_{\max})^{-1/2}\sum_{t=h_{\min}}^{n-1}E\max_{h\in H_n}\left\|\left(u_{t+1,h}-u_{t+1,h_{\max}}\right)Y_{t,h}'\right\| &\leq \frac{(n-h_{\min})}{n^{1/2}(n-h_{\max})^{1/2}}n^{1/2}c\sum_{j\geq h_{\min}+1}\|\pi_j\|\,o(n^{1/6}) \\ &= o(n^{1/6}).\end{aligned}$$

Also, $\max_h\left\|l(h)'\left(\left(\hat{\Gamma}_h^{-1}-\Gamma_h^{-1}\right)\otimes I_p\right)\right\| \leq pM_2^{1/2}\max_h\left\|\hat{\Gamma}_h^{-1}-\Gamma_h^{-1}\right\|_2 = o_p(n^{-1/3+\delta/2})$ such that $w_{51n} = o_p(1)$.

Use the notation $l(h) = \left(l_{1,h}',...,l_{h,h}'\right)'$ where $l_{j,h}$ is a $p^2\times 1$ vector with $l_{j,h} = l_j$ for part i) and $\Gamma_h^{jk}$ is the $j,k$-th block of $\Gamma_h^{-1}$ and note that

$$\begin{aligned}E\max_{h\in H_n}|w_{52n}(h)|^2 &\leq \sum_{h=h_{\min}}^{h_{\max}}E|w_{52n}(h)|^2 \\ &\leq \sum_{h=h_{\min}}^{h_{\max}}(n-h_{\max})^{-1}E\left(\sum_{t=h}^{n-1}\sum_{k_1,k_2=1}^{h}\sum_{j=h+1}^{h_{\max}}l_{k_1,h}'\left(\Gamma_h^{k_1,k_2}y_{t-k_2}\otimes\pi_j y_{t-j}\right)\right)^2.\end{aligned}$$

From Corollary 4.3 it follows that

$$\begin{aligned}&E\left(\sum_{t=h}^{n-1}\sum_{k_1,k_2=1}^{h}\sum_{j=h+1}^{h_{\max}}l_{k_1,h}'\left(\Gamma_h^{k_1,k_2}y_{t-k_2}\otimes\pi_j y_{t-j}\right)\right)^2 \\ =\;&\sum_{t,s=h}^{n-1}\sum_{k_1,...,k_4=1}^{h}\sum_{j_1,j_2=h+1}^{h_{\max}}l_{k_1,h}'\,\text{vec}\left(\pi_{j_1}\Gamma_{k_2-j_1}^{yy}\left(\Gamma_h^{k_1,k_2}\right)'\right)\text{vec}\left(\pi_{j_2}\Gamma_{k_3-j_2}^{yy}\left(\Gamma_h^{k_3,k_4}\right)'\right)'l_{k_4,h} \\ &+\sum_{t,s=h}^{n-1}\sum_{k_1,...,k_4=1}^{h}\sum_{j_1,j_2=h+1}^{h_{\max}}l_{k_1,h}'\left(\Gamma_h^{k_1,k_2}\Gamma_{t-s+j_2-k_2}^{yy}\pi_{j_2}\otimes\pi_{j_1}\Gamma_{t-s+k_3-j_1}^{yy}\left(\Gamma_h^{k_3,k_4}\right)'\right)K_{pp}l_{k_4,h} \\ &+\sum_{t,s=h}^{n-1}\sum_{k_1,...,k_4=1}^{h}\sum_{j_1,j_2=h+1}^{h_{\max}}l_{k_1,h}'\left(\Gamma_h^{k_1,k_2}\Gamma_{t-s+k_3-k_2}^{yy}\Gamma_h^{k_3,k_4}\otimes\pi_{j_1}\Gamma_{t-s+j_2-j_1}^{yy}\pi_{j_2}'\right)l_{k_4,h} \\ &+\sum_{t,s=h}^{n-1}\sum_{k_1,...,k_4=1}^{h}\sum_{j_1,j_2=h+1}^{h_{\max}}l_{k_1,h}'\left(\Gamma_h^{k_1,k_2}\otimes\pi_{j_1}\right)\mathcal{K}_4(t-k_2,t-j_1,s-k_3,s-j_2)\left(\Gamma_h^{k_3,k_4}\otimes\pi_{j_2}'\right)l_{k_4,h}\end{aligned}$$

For the first term note that

$$\begin{aligned}
&\textstyle\sum_{t,s=h}^{n-1}\left(\sum_{j=h+1}^{h_{\max}}\sum_{k_1,k_2=1}^{h} l_{k_1,h}'\operatorname{vec}\left(\pi_j\Gamma_{k_2-j}^{yy}\left(\Gamma_h^{k_1,k_2}\right)'\right)\right)^2\\
\leq\;& n^2p\left(\textstyle\sum_{j=h+1}^{h_{\max}}\|\pi_j\|^2\right)\left(\sum_{j=h+1}^{h_{\max}}\left\|\sum_{k_1,k_2=1}^{h} l_{k_1,h}'\left(\Gamma_h^{k_1,k_2}\Gamma_{j-k_2}^{yy}\otimes I_p\right)\right\|^2\right)\\
\leq\;& Kn^2p\left(\textstyle\sum_{j=h_{\min}+1}^{\infty}\|\pi_j\|\right)^2
\end{aligned}$$

because

$$\textstyle\sum_{j=h+1}^{h_{\max}}\left\|\sum_{k_1,k_2=1}^{h} l_{k_1,h}'\left(\Gamma_h^{k_1,k_2}\Gamma_{j-k_2}^{yy}\otimes I_p\right)\right\|^2\leq\|l(h)\|^2\left\|\Gamma_h^{-1}\Gamma_h\right\|_2^2<K$$

where $\Gamma_h$ is a matrix with $k,l$-th element $\Gamma_{l+h-k+1}^{yy}$ and $K$ is a generic bounded constant that does not depend on $h$. Then

$$\begin{aligned}
&\textstyle\sum_{h=h_{\min}}^{h_{\max}}\left(n-h_{\max}\right)^{-1}Kn^2p\left(\sum_{j=h_{\min}+1}^{\infty}\|\pi_j\|\right)^2\\
=\;& Knp\left(n-h_{\max}\right)^{-1}\Delta_n\left(n^{1/2}\textstyle\sum_{j=h_{\min}+1}^{\infty}\|\pi_j\|\right)^2\\
\leq\;& Knp\left(n-h_{\max}\right)^{-1}\left(n^{1/2}\textstyle\sum_{j=h_{\min}+1}^{\infty}\|\pi_j\|\right)\to 0
\end{aligned}$$

where the inequality follows from (4.1). For the second term consider the following term of equal order

$$\textstyle\sum_{t,s=h}^{n-1}\left\|\sum_{j=h+1}^{h_{\max}}\sum_{k_1,k_2=1}^{h} l_{k_1}'\left(\Gamma_h^{k_1,k_2}\Gamma_{t-s+j-k_2}^{yy}\pi_j\otimes I_p\right)\right\|^2\leq Knh_{\max}^2\left(\sum_{j=h_{\min}+1}^{\infty}\|\pi_j\|\right)^2$$

which only differs by $K_{pp}$. The inequality holds because

$$\begin{aligned}
&\textstyle\sum_{t,s=h}^{n-1}\left(\sum_{j=1}^{h_{\max}}\left\|\sum_{k_1,k_2=1}^{h} l_{k_1}'\left(\Gamma_h^{k_1,k_2}\Gamma_{t-k_2-s+j}^{yy}\otimes I_p\right)\right\|^2\right)\left(\sum_{j=h_{\min}+1}^{\infty}\|\pi_j\|^2\right)\\
\leq\;& \textstyle\sum_{t,s=h}^{n-1}\sum_{j=1}^{h_{\max}}\sum_{k_2=1}^{h}\left\|\Gamma_{t-k_2-s+j}^{yy}\right\|^2\left(\sum_{k_2=1}^{h}\left\|\sum_{k_1=1}^{h} l_{k_1,h}'\left(\Gamma_h^{k_1,k_2}\otimes I_p\right)\right\|^2\right)\left(\sum_{j=h_{\min}+1}^{\infty}\|\pi_j\|\right)^2\\
\leq\;& \textstyle\sum_{j=1}^{h_{\max}}\sum_{k_2=1}^{h_{\max}}\sum_{u=-n+1}^{n-1}\left(n-|u|\right)\left\|\Gamma_{u-k_2+j}^{yy}\right\|^2\left(\sum_{k_2=1}^{h}\left\|\sum_{k_1=1}^{h} l_{k_1,h}'\left(\Gamma_h^{k_1,k_2}\otimes I_p\right)\right\|^2\right)\left(\sum_{j=h_{\min}+1}^{\infty}\|\pi_j\|\right)^2\\
=\;& O\left(h_{\max}^2 n\left(\textstyle\sum_{j=h_{\min}+1}^{\infty}\|\pi_j\|\right)^2\right)
\end{aligned}$$

such that the second term is of smaller order than the first term. Note that here

$$\textstyle\sum_{k_2=1}^{h}\left\|\sum_{k_1=1}^{h} l_{k_1}'\left(\Gamma_h^{k_1,k_2}\otimes I_p\right)\right\|^2=O(1)$$

because $\left\|l(h)'\left(\Gamma_h^{-1}\otimes I_p\right)\right\|\leq\|l(h)\|\left\|\Gamma_h^{-1}\right\|_2$ is uniformly bounded in $h$. Finally, turning to the third term,

$$\sum_{k_1,\dots,k_4=1}^{h}\left|l'_{k_1,h}\left(\Gamma_h^{k_1,k_2}\otimes I_p\right)\sum_{t,s=h}^{n-1}\left(\Gamma^{yy}_{t-s+k_3-k_2}\otimes\sum_{j_1,j_2=h+1}^{h_{\max}}\pi_{j_1}\Gamma^{yy}_{t-s+j_2-j_1}\pi'_{j_2}\right)\left(\Gamma_h^{k_3,k_4}\otimes I_p\right)l_{k_4,h}\right|$$

$$\leq \left(\sum_{k_2=1}^{h}\left\|\sum_{k_1=1}^{h}l'_{k_1}\left(\Gamma_h^{k_1,k_2}\otimes I_p\right)\right\|^2\right)^{1/2}\left(\sum_{k_3=1}^{h}\left\|\sum_{k_4=1}^{h}\left(\Gamma_h^{k_3,k_4}\otimes I_p\right)l_{k_4}\right\|^2\right)^{1/2}$$

$$\times\left(\sum_{k_2=1}^{h}\sum_{k_3=1}^{h}\left\|\sum_{t,s=h}^{n-1}\left(\Gamma^{yy}_{t-s+k_3-k_2}\otimes\sum_{j_1,j_2=h+1}^{h_{\max}}\pi_{j_1}\Gamma^{yy}_{t-s+j_2-j_1}\pi'_{j_2}\right)\right\|^2\right)^{1/2}$$

$$= O\left(h_{\max}n\left(\sum_{j=h_{\min}+1}^{\infty}\|\pi_j\|\right)^2\right)$$

such that the third term is also of smaller order than the first. Finally, the fourth order cumulant term is of smaller order by Corollary 4.3. Therefore, $w_{52n}=o_p(1)$.

For $|w_{6n}|$ note

$$|w_{6n}|\leq\max_{h\in H_n}\left\|l(h)'\left(\hat{\Gamma}_h^{-1}\otimes I_p\right)\right\|(n-h_{\max})^{-1/2}\sum_{t=h_{\min}}^{h_{\max}}\|u_{t+1,h_{\max}}\|\left\|Y'_{t,h}\right\| \tag{4.7}$$

where $\|u_{t+1,h_{\max}}\|\left\|Y'_{t,h}\right\|\leq\left(\|y_{t+1}\|+\sum_{j=1}^{\infty}\|\pi_j\|\|y_{t+1-j}\|\right)\max_{h\in H_n}\left\|Y'_{t,h}\right\|$ and

$$E\left(\|y_{t+1-j}\|\max_{h\in H_n}\left\|Y'_{t,h}\right\|\right)=o(n^{1/6})$$

by previous arguments such that the last term in (4.7) is bounded in expectation by

$$O\left(\Delta_n h_{\max}^{1/2}(n-h_{\max})^{-1/2}\right)=O\left(n^{-1/3+\delta}\right)$$

and thus $|w_{6n}|=o_p(1)$.

Finally, consider $|w_{7n}|$. We distinguish the following terms

$$\begin{aligned}
w_{7n} &= l(h)'\left(\left(\hat{\Gamma}_h^{-1}-\Gamma_h^{-1}\right)\otimes I_p\right)\operatorname{vec}\left((n-h_{\max})^{-1/2}\left(U_{1n}(h)+U_{2n}(h)\right)\right)\\
&\quad -l(h_{\max})'\left(\left(\hat{\Gamma}_{h_{\max}}^{-1}-\Gamma_{h_{\max}}^{-1}\right)\otimes I_p\right)\operatorname{vec}\left((n-h_{\max})^{-1/2}\left(U_{1n}(h_{\max})+U_{2n}(h_{\max})\right)\right)\\
&\quad +(n-h_{\max})^{-1/2}\sum_{t=h_{\max}}^{n-1}\left(l(h)'\left(\Gamma_h^{-1}Y_{t,h}\otimes I_p\right)-l(h_{\max})'\left(\Gamma_{h_{\max}}^{-1}Y_{t,h_{\max}}\otimes I_p\right)\right)u_{t+1,h_{\max}}\\
&= w_{71n}-w_{72n}+w_{73n}
\end{aligned}$$

where $w_{71n},...,w_{73n}$ are defined in the obvious way and $U_{1n}(h)=\sum_{t=h_{\max}}^{n-1}v_{t+1}Y'_{t,h}$ and

$$U_{2n}(h)=\sum_{t=h_{\max}}^{n-1}\left(u_{t+1,h_{\max}}-v_{t+1}\right)Y'_{t,h}.$$

For the term $w_{72n}$ the proof of Theorem 2 in Lewis and Reinsel (1985) can be applied to show that $w_{72n}=o_p(1)$. For $w_{71n}$ and $w_{73n}$ we need additional uniformity arguments. For $w_{71}$ consider

$$l(h)'\left(\left(\hat{\Gamma}_h^{-1}-\Gamma_h^{-1}\right)\otimes I_p\right)=l(h)'\left(\Gamma_h^{-1}\otimes I_p\right)\left(\left(\Gamma_h-\hat{\Gamma}_h\right)\otimes I_p\right)\left(\hat{\Gamma}_h^{-1}\otimes I_p\right)$$

where the vector $a(h) \equiv l(h)' \left(\Gamma_h^{-1} \otimes I_p\right)$ satisfies $\|a(h)\|^2 < \infty$ for all $h$. Then

$$|w_{71n}| \leq \max_{h \in H_n} \left\| a(h)' \left( \left( \Gamma_h - \hat{\Gamma}_h \right) \otimes I_p \right) \right\| \max_{h \in H_n} \left\| \hat{\Gamma}_h^{-1} \right\|_2 \max_{h \in H_n} \left\| (n - h_{\max})^{-1/2} \left( U_{1,n}(h) + U_{2,n}(h) \right) \right\|$$

Using the result in (4.6) it follows that $\max_{h \in H_n} \left\| \hat{\Gamma}_h^{-1} \right\|_2 = O_p(1)$. Next consider

$$\max_{h \in H_n} \left\| a(h)' \left( \left( \Gamma_h - \hat{\Gamma}_h \right) \otimes I_p \right) \right\|^2 \leq p \max_{h \in H_n} \|a(h)\|^2 \max_{h \in H_n} \left\| \Gamma_h - \hat{\Gamma}_h \right\|^2$$

such that it follows from Lemma 4.4 that

$$E \max_{h \in H_n} \left\| a(h)' \left( \left( \Gamma_h - \hat{\Gamma}_h \right) \otimes I_p \right) \right\|^2 = O(n^{-2/3+\delta}).$$

Next, consider

$$E \max_{h \in H_n} \left\| (n - h_{\max})^{-1/2} U_{1,n}(h) \right\|^2 \leq (n - h_{\max})^{-1} \sum_{t=h_{\max}}^{n-1} E \left( v_{t+1}' v_{t+1} \right) E Y_{t,h_{\max}}' Y_{t,h_{\max}} = o(n^{1/3}).$$

Finally,

$$E \max_{h \in H_n} \left\| (n - h_{\max})^{-1/2} U_{2,n}(h) \right\| \leq \left( E \left\| Y_{t,h_{\max}}' \right\|^2 \right)^{1/2} \frac{n - h_{\max}}{(n - h_{\max})^{1/2}} c \sum_{j=h_{\max}+1}^{\infty} \|\pi_j\| = o(n^{1/6})$$

with $\sup_t \left( E \left\| y_t^2 \right\| \right)^{1/2} \leq c < \infty$. This shows $\max_{h \in H_n} |w_{71n}| = o_p(n^{-1/3+\delta/2} n^{1/6}) = o_p(1)$.

Next, let $\zeta_{t,h} = l(h)' \left( \Gamma_h^{-1} Y_{t,h} \otimes I_p \right)$ such that

$$w_{73n} = (n - h_{\max})^{-1/2} \textstyle\sum_{t=h_{\max}}^{n-1} \left( \zeta_{t,h} - \zeta_{t,h_{\max}} \right) \left( u_{t+1,h_{\max}} - v_{t+1} + v_{t+1} \right)$$

and $E \left\| \zeta_{t,h} \right\|^2 = l(h)' \left( \Gamma_h^{-1} \otimes I_p \right) l(h) \leq C < \infty$ by Lewis and Reinsel (1985, p.399). Then

$$\begin{aligned} & (n - h_{\max})^{-1/2} \textstyle\sum_{t=h_{\max}}^{n-1} E \left\| \left( \zeta_{t,h} - \zeta_{t,h_{\max}} \right) \left( u_{t+1,h_{\max}} - v_{t+1} \right) \right\| \\ \leq \; & \left( E \max_{h \in H_n} \left\| \left( \zeta_{t,h} - \zeta_{t,h_{\max}} \right) \right\|^2 \right)^{1/2} \left( E \|y_t\|^2 \right)^{1/2} (n - h_{\max})^{1/2} \sum_{j=h_{\max}+1}^{\infty} \|\pi_j\| \to 0 \end{aligned}$$

where $E \max_{h \in H_n} \left\| \left( \zeta_{t,h} - \zeta_{t,h_{\max}} \right) \right\|^2 = o(1)$ by the analysis below and

$$\begin{aligned} & E \left\| (n - h_{\max})^{-1/2} \textstyle\sum_{t=h_{\max}}^{n-1} \left( \zeta_{t,h} - \zeta_{t,h_{\max}} \right) v_{t+1} \right\|^2 \\ = \; & (n - h_{\max})^{-1} \textstyle\sum_{s,t=h_{\max}}^{n-1} E \left[ \left( \zeta_{t,h} - \zeta_{t,h_{\max}} \right) v_{t+1} v_{s+1}' \left( \zeta_{s,h} - \zeta_{s,h_{\max}} \right)' \right] \\ = \; & E \left[ \left( \zeta_{t,h} - \zeta_{t,h_{\max}} \right) \Sigma_v \left( \zeta_{t,h} - \zeta_{t,h_{\max}} \right)' \right] \\ \leq \; & \|\Sigma_v\| \, E \left\| \left( \zeta_{t,h} - \zeta_{t,h_{\max}} \right) \right\|^2 \end{aligned}$$

where we have used stationarity and the fact that $Ev_{t+1}v'_{s+1} = E(v_{t+1}v'_{s+1}|\mathcal{F}_t) = 0$ for $t > s$ and $E(v_{t+1}v'_{t+1}|\mathcal{F}_t) = \Sigma_v$.

At this point the proof for Theorem 2.3 i) and ii) proceeds separately. First turn to i). Since $h \leq h_{\max}$,

$$\begin{aligned} \left\|\zeta_{t,h} - \zeta_{t,h_{\max}}\right\| &= \left\|\sum_{j,l=1}^{h} l'_j \left(\Gamma_h^{j,l} y_{t-l} \otimes I_p\right) - \sum_{j,l=1}^{h_{\max}} l'_j \left(\Gamma_{h_{\max}}^{j,l} y_{t-l} \otimes I_p\right)\right\| \\ &\leq \sum_{j,l=1}^{h} \left\|l'_j \left(\left(\Gamma_h^{j,l} - \Gamma_{h_{\max}}^{j,l}\right) y_{t-l} \otimes I_p\right)\right\| + \sum_{j=1,l=h_{\min}+1}^{h_{\max},h_{\max}} \left\|l'_j \left(\Gamma_{h_{\max}}^{j,l} \otimes I_p\right)\right\| \left\|y_{t-l}\right\| \\ &\quad + \sum_{j=h_{\min}+1,l=1}^{h_{\max},h_{\max}} \left\|l'_j \left(\Gamma_{h_{\max}}^{j,l} \otimes I_p\right)\right\| \left\|y_{t-l}\right\| \end{aligned}$$

where $\sup_t E\left\|y_t\right\|^2 \leq c < \infty$. By Hannan and Deistler (1988, Theorem 6.6.11) there exists a constant $c_1$ such that $\sup_h \sup_{j \leq h} \sum_{l=1}^{h} \left\|\Gamma_h^{j,l}\right\| \leq c_1/p < \infty$. By the second part of Lemma (4.7) and for any $\varepsilon > 0$ there exists a constant $k_0 < \infty$ such that $\sup_{j \leq k_0} \sum_{l=h_{\min}+1}^{h_{\max}} \left\|\Gamma_{h_{\max}}^{j,l}\right\| \to 0$ and $\sum_{j=k_0+1}^{\infty} \left\|l_j\right\| < \varepsilon c_1^{-1}$. Then

$$\begin{aligned} \sum_{j=1,l=h_{\min}+1}^{h_{\max},h_{\max}} \left\|l'_j \left(\Gamma_{h_{\max}}^{j,l} \otimes I_p\right)\right\| &\leq p \sup_{j \leq k_0} \sum_{l=h_{\min}+1}^{h_{\max}} \left\|\Gamma_{h_{\max}}^{j,l}\right\| \sum_{j=1}^{k_0} \left\|l_j\right\| \\ &\quad + p \sup_{j \leq h_{\max}} \sum_{l=h_{\min}+1}^{h_{\max}} \left\|\Gamma_{h_{\max}}^{j,l}\right\| \sum_{j=k_0+1}^{\infty} \left\|l_j\right\| \\ &\leq o(1) + \varepsilon. \end{aligned}$$

Also

$$\sum_{j=h_{\min}+1,l=1}^{h_{\max},h_{\max}} \left\|l'_j\right\| \left\|\Gamma_{h_{\max}}^{j,l}\right\| \leq \sup_{j \leq h_{\max}} \sum_{l=1}^{h_{\max}} \left\|\Gamma_{h_{\max}}^{j,l}\right\| \sum_{j=h_{\min}+1}^{h_{\max}} \left\|l_j\right\| \to 0$$

by the assumptions on $l$. Now define constants $c_1 = 2\sum_{j=1}^{\infty} \left\|l_j\right\|$ and $c_2 = 4 \sup_{h \in H_n} \sup_{j \leq h} \sum_{l=1}^{h} \left\|\Gamma_h^{j,l}\right\|$. For any $\varepsilon > 0$ fix integer constants $k_0, k_1$ such that $\sum_{j=k_0+1}^{\infty} \left\|l_j\right\| < \varepsilon c_2^{-1}$ and

$$\sup_{h \in H_n} \sup_{j \leq k_0} \sum_{l=k_1+1}^{h} \left(\left\|\Gamma_h^{j,l}\right\| + \left\|\Gamma_{h_{\max}}^{j,l}\right\|\right) < \varepsilon c_1^{-1}$$

where the last inequality holds for some $k_1$ and any $k_0$ and all $n \geq n_0$ for some positive integer $n_0 < \infty$ by Lemma (4.7). Then

$$\begin{aligned} \sup_{h \in H_n} \sum_{j,l=1}^{h} \left\|l_j\right\| \left\|\Gamma_h^{j,l} - \Gamma_{h_{\max}}^{j,l}\right\| &\leq \sum_{j,l=1}^{k_0,k_1} \left\|l_j\right\| \sup_{h \in H_n} \left\|\Gamma_h^{j,l} - \Gamma_{h_{\max}}^{j,l}\right\| \\ &\quad + \sum_{j=k_0+1}^{\infty} \left\|l_j\right\| \sup_{h \in H_n} \sup_{j \leq h} \sum_{l=1}^{h} \left(\left\|\Gamma_h^{j,l}\right\| + \left\|\Gamma_{h_{\max}}^{j,l}\right\|\right) \\ &\quad + \sum_{j=1}^{k_0} \left\|l_j\right\| \sup_{h \in H_n} \sup_{j \leq k_0} \sum_{l=k_1+1}^{h} \left(\left\|\Gamma_h^{j,l}\right\| + \left\|\Gamma_{h_{\max}}^{j,l}\right\|\right) \\ &\leq o(1) + \varepsilon \end{aligned}$$

because for fixed $k_0, k_1$, $\sup_{h \in H_n} \left\|\Gamma_h^{j,l} - \Gamma_{h_{\max}}^{j,l}\right\| \to 0$ by Lewis and Reinsel (1985, p. 402) and Hannan and Deistler (1988, Theorem 6.6.12) such that $\max_{h \in H_n} \left\|\zeta_{t,h} - \zeta_{t,h_{\max}}\right\| = o_p(1)$.

Now turn to the proof for Theorem 2.3 ii). Partition $l(h_{\max}) = (\tilde{l}(h_{\max})', \tilde{l}_1(h_{\max})')'$ where $\tilde{l}(h_{\max}) = \left(l'_{1,h_{\max}}, ..., l'_{h,h_{\max}}\right)'$, $\tilde{l}_1(h_{\max})' = \left(l'_{h+1,h_{\max}}, ..., l'_{h_{\max},h_{\max}}\right)'$. Consider

$$E \max_{h \in H_n} \left\|\zeta_{t,h} - \zeta_{t,h_{\max}}\right\|^2 \leq \sum_{h=h_{\min}}^{h_{\max}} E \left\|\zeta_{t,h} - \zeta_{t,h_{\max}}\right\|^2$$

with

$$
\begin{aligned}
(4.8) \qquad & E \left\|\zeta_{t,h} - \zeta_{t,h_{\max}}\right\|^2 \\
= \;& l(h)' \left(\Gamma_h^{-1} \otimes I_p\right) l(h) - 2l(h)' \left(\Gamma_h^{-1} \otimes I_p\right) \tilde{l}(h_{\max}) + l(h_{\max})' \left(\Gamma_{h_{\max}}^{-1} \otimes I_p\right) l(h_{\max}) \\
\leq \;& \left|l(h)' \left(\Gamma_h^{-1} \otimes I_p\right) \left(l(h) - \tilde{l}(h_{\max})\right)\right| + \left|\left(l(h) - \tilde{l}(h_{\max})\right)' \left(\Gamma_h^{-1} \otimes I_p\right) \tilde{l}(h_{\max})\right| \\
& + \left|\tilde{l}(h_{\max})' \left(\Gamma_h^{-1} \otimes I_p\right) \tilde{l}(h_{\max}) - l(h_{\max})' \left(\Gamma_{h_{\max}}^{-1} \otimes I_p\right) l(h_{\max})\right|.
\end{aligned}
$$

Note that for any sequence $h^{**} \to \infty$ such that $\sum_{j_1=1}^{h^{**}} \|l_{j_1,h} - l_{j_2,h_{\max}}\|^2 = o(\Delta_n^{-2})$ and $\Delta_n/\sqrt{h^{**}} \to 0$,

$$
\begin{aligned}
& \left|\left(l(h) - \tilde{l}(h_{\max})\right)' \left(\Gamma_h^{-1} \otimes I_p\right) \tilde{l}(h_{\max})\right| \\
\leq \;& \left(\textstyle\sum_{j_1=1}^{h^{**}} \|l_{j_1,h} - l_{j_1,h_{\max}}\|^2\right)^{1/2} \left(\textstyle\sum_{j_1=1}^{h^{**}} \left\|\sum_{j_2=1}^{h} \left(\Gamma_h^{j_1,j_2} \otimes I_p\right) l_{j_2,h_{\max}}\right\|^2\right)^{1/2} \\
& + \left(\textstyle\sum_{j_1=h^{**}+1}^{h} (j_1)^2 \|l_{j_1,h} - l_{j_1,h_{\max}}\|^2\right)^{1/2} \left(\textstyle\sum_{j_1=h^{**}+1}^{h} (j_1)^{-2} \left\|\sum_{j_2=1}^{h} \left(\Gamma_h^{j_1,j_2} \otimes I_p\right) l_{j_2,h_{\max}}\right\|^2\right)^{1/2} \\
= \;& o(\Delta_n^{-1})
\end{aligned}
$$

since $\sum_{j_1=1}^{h^{**}} \left\|\sum_{j_2=1}^{h} \left(\Gamma_h^{j_1,j_2} \otimes I_p\right) l_{j_2,h_{\max}}\right\|^2 \leq \left\|\left(\Gamma_h^{-1} \otimes I_p\right) \tilde{l}(h_{\max})\right\|^2$, $\left\|\Gamma_h^{-1}\right\|_2$ is uniformly bounded by Lewis and Reinsel (1985, p.397) and $\left\|\tilde{l}(h_{\max})\right\| \leq \sqrt{M_2}$ uniformly in $h$. Since also

$$\textstyle\sum_{j_1=h^{**}+1}^{h} j_1^2 \|l_{j_1,h} - l_{j_2,h_{\max}}\|^2 < \infty$$

and

$$\left(\textstyle\sum_{j_1=h^{**}+1}^{h} j_1^{-2} \left\|\sum_{j_2=1}^{h} \left(\Gamma_h^{j_1,j_2} \otimes I_p\right) l_{j_2,h_{\max}}\right\|^2\right)^{1/2} = O\left(\left(\textstyle\sum_{j_1=h^{**}+1}^{h} j_1^{-2}\right)^{1/2}\right) = O\left(1/\sqrt{h^{**}}\right)$$

the second term is $o(\Delta_n^{-1})$ as well. Finally, consider

$$
\begin{aligned}
& \left|\tilde{l}(h_{\max})' \left(\Gamma_h^{-1} \otimes I_p\right) \tilde{l}(h_{\max}) - l(h_{\max})' \left(\Gamma_{h_{\max}}^{-1} \otimes I_p\right) l(h_{\max})\right| \\
\leq \;& \left|\tilde{l}(h_{\max})' \left(\left(\Gamma_h^{-1} - \left[\Gamma_{h_{\max}}^{-1}\right]_{11}\right) \otimes I_p\right) \tilde{l}(h_{\max})\right| + 2\left|\tilde{l}(h_{\max})' \left(\left[\Gamma_{h_{\max}}^{-1}\right]_{12} \otimes I_p\right) \tilde{l}_1(h_{\max})\right| \\
& + \left|\tilde{l}_1(h_{\max})' \left(\left[\Gamma_{h_{\max}}^{-1}\right]_{22} \otimes I_p\right) \tilde{l}_1(h_{\max})\right|
\end{aligned}
$$

where $\Gamma_{h_{\max}}^{-1}$ and $\Gamma_{h_{\max}}$ are partitioned as in Lemma (4.8) where the notation for the blocks of the inverse $\left[\Gamma_{h_{\max}}^{-1}\right]_{ij}$ and the blocks $\Gamma_{ij,h_{\max}}$ of $\Gamma_{h_{\max}}$ is introduced. Then, $\left\|\tilde{l}_1(h_{\max})\right\| = o(\Delta_n^{-1})$ uniformly in $h$ because by assumption $\exists \underline{h} \leq h_{\min}$ such that $\sum_{i=\underline{h}+1}^{h_{\max}} \|l_{i,h_{\max}}\|^2 = o(\Delta_n^{-2})$. Then,

$$\left|\tilde{l}(h_{\max})' \left(\left[\Gamma_{h_{\max}}^{-1}\right]_{12} \otimes I_p\right) \tilde{l}_1(h_{\max})\right| = o(\Delta_n^{-1})$$

and

$$\left|\tilde{l}_1(h_{\max})' \left(\left[\Gamma_{h_{\max}}^{-1}\right]_{22} \otimes I_p\right) \tilde{l}_1(h_{\max})\right| = o(\Delta_n^{-1}).$$

For the first term define $a(h) = \tilde{l}(h_{\max})' \left(\Gamma_h^{-1} \otimes I_p\right)$ and $\tilde{a}(h) = \tilde{l}(h_{\max})' \left(\left[\Gamma_{h_{\max}}^{-1}\right]_{11} \otimes I_p\right)$ where $\|a(h)\| < \infty$ and $\|\tilde{a}(h)\| < \infty$ uniformly in $h$. Let $A$ be defined as in Lemma (4.8). Write $\Gamma_h^{-1} - \left[\Gamma_{h_{\max}}^{-1}\right]_{11} = -\Gamma_h^{-1} A \left[\Gamma_{h_{\max}}^{-1}\right]_{11}$ by the partitioned inverse formula. Now for $h^* = \underline{h} + (h_{\min} - \underline{h})/2$ such that $\underline{h} \leq h^* \leq h_{\min}$ with $h^* - \underline{h} = (h_{\min} - \underline{h})/2$ and $h_{\min} - h^* = (h_{\min} - \underline{h})/2$ it follows that

$$\begin{aligned}
\left|a(h)'(A \otimes I_p)\tilde{a}(h)\right| &= \left|\textstyle\sum_{i_1=1}^{h} l'_{i_1,h} \sum_{j_1,j_2=1}^{h} \sum_{i_2=1}^{h} \left(\Gamma_h^{i_1,j_1} A_{j_1,j_2} \Gamma_{h_{\max}}^{j_2,i_2} \otimes I_p\right) l_{i_2,h_{\max}}\right| \\
&\leq \left|\textstyle\sum_{i_1=1}^{\underline{h}} l'_{i_1,h} \sum_{j_1,j_2=1}^{h} \sum_{i_2=1}^{h} \left(\Gamma_h^{i_1,j_1} A_{j_1,j_2} \Gamma_{h_{\max}}^{j_2,i_2} \otimes I_p\right) l_{i_2,h_{\max}}\right| \\
&\quad + \left(\sum_{i_1=\underline{h}+1}^{h} \left\|l'_{i_1,h}\right\|^2\right)^{1/2} \left(\sum_{i_1=\underline{h}+1}^{h} \left\|\sum_{j_1,j_2=1}^{h} \sum_{i_2=1}^{h} \left(\Gamma_h^{i_1,j_1} A_{j_1,j_2} \Gamma_{h_{\max}}^{j_2,i_2} \otimes I_p\right) l_{i_2,h_{\max}}\right\|^2\right)^{1/2}
\end{aligned}$$

where the second term is $o(\Delta_n^{-1})$ uniformly in $h$ by the assumptions on the sequence $l(h)$ and the fact that $\|(A \otimes I_p)\tilde{a}(h)\|$ is uniformly bounded in $h$. For the first term consider

$$\begin{aligned}
&\left|\textstyle\sum_{i_1=1}^{\underline{h}} l'_{i_1,h} \sum_{j_1,j_2=1}^{h} \sum_{i_2=1}^{h} \left(\Gamma_h^{i_1,j_1} A_{j_1,j_2} \Gamma_{h_{\max}}^{j_2,i_2} \otimes I_p\right) l_{i_2,h_{\max}}\right| \\
= &\left|\textstyle\sum_{i_1,i_2=1}^{\underline{h}} \sum_{j_1,j_2=1}^{h} l'_{i_1,h} \left(\Gamma_h^{i_1,j_1} A_{j_1,j_2} \Gamma_{h_{\max}}^{j_2,i_2} \otimes I_p\right) l_{i_2,h_{\max}}\right| + o(\Delta_n^{-1})
\end{aligned}$$

where the order of the error term follows again from $\left(\sum_{i_1=\underline{h}+1}^{h} \left\|l'_{i_1,h}\right\|^2\right)^{1/2} = o\left(\Delta_n^{-1}\right)$ and

$$\begin{aligned}
&\left|\textstyle\sum_{i_1,i_2=1}^{\underline{h}} \sum_{j_1,j_2=1}^{h} l'_{i_1,h} \left(\Gamma_h^{i_1,j_1} A_{j_1,j_2} \Gamma_{h_{\max}}^{j_2,i_2} \otimes I_p\right) l_{i_2,h_{\max}}\right| \\
\leq\ & p \textstyle\sum_{j_1=1}^{h^*} \left\|\sum_{i_1=1}^{\underline{h}} l'_{i_1,h} \left(\Gamma_h^{i_1,j_1} \otimes I_p\right)\right\| \sum_{j_2=1}^{h} \|A_{j_1,j_2}\| \left\|\sum_{i_2=1}^{\underline{h}} \left(\Gamma_{h_{\max}}^{j_2,i_2} \otimes I_p\right) l_{i_2,h_{\max}}\right\| \\
&+ \left(\textstyle\sum_{j_1=h^*}^{h} \sum_{i_1=1}^{\underline{h}} \left\|\Gamma_h^{i_1,j_1}\right\|^2\right)^{1/2} \left(\textstyle\sum_{i_1=1}^{\underline{h}} \left\|l'_{i_1,h}\right\|^2\right)^{1/2} \\
&\times \left(\textstyle\sum_{j_1=h^*}^{h} \left\|\sum_{j_2=1}^{h} \sum_{i_2=1}^{\underline{h}} \left(A_{j_1,j_2} \Gamma_{h_{\max}}^{j_2,i_2} \otimes I_p\right) l_{i_2,h_{\max}}\right\|^2\right)^{1/2} \\
\leq\ & c_1 \textstyle\sum_{j_1=1}^{h^*} \sum_{j_2=1}^{h} \|A_{j_1,j_2}\| \left\|\sum_{i_1=1}^{\underline{h}} \left(\Gamma_{h_{\max}}^{j_2,i_2} \otimes I_p\right) l_{i_2,h_{\max}}\right\| \\
&+ \left(\textstyle\sum_{j_1=h^*}^{h} \sum_{i_1=1}^{\underline{h}} \left\|\Gamma_h^{i_1,j_1}\right\|^2\right)^{1/2} \|l(h)\| \left\|\left(A \left[\Gamma_{h_{\max}}^{-1}\right]_{11} \otimes I_p\right) \tilde{l}(h_{\max})\right\|
\end{aligned}$$

where $\left\|\sum_{i_1=1}^{\underline{h}} l'_{i_1,h}\left(\Gamma_h^{i_1,j_1}\otimes I_p\right)\right\| < c_1$ is uniformly bounded, $\sum_{j_1=h^*}^{h}\sum_{i_1=1}^{\underline{h}}\left\|\Gamma_h^{i_1,j_1}\right\|^2 = o\left((h_{\min}-\underline{h})^{-1}\right)$ by Lemma (4.7) and $\sum_{j_1=1}^{h^*}\sum_{j_2=1}^{h}\|A_{j_1,j_2}\|\left\|\sum_{i_2=1}^{\underline{h}}\left(\Gamma_{h_{\max}}^{j_2,i_2}\otimes I_p\right)l_{i_2,h_{\max}}\right\| = o\left((h_{\min}-\underline{h})^{-1}\right)$ because $\left\|\sum_{i_2=1}^{\underline{h}}\left(\Gamma_{h_{\max}}^{j_2,i_2}\otimes I_p\right)l_{i_2,h_{\max}}\right\|$ is uniformly bounded and $\sum_{j_1=1}^{h^*}\sum_{j_2=1}^{h}\|A_{j_1,j_2}\| = o\left((h_{\min}-\underline{h})^{-1}\right)$ by Lemma (4.8). It now follows that $\sum_{h=h_{\min}}^{h_{\max}} E\left\|\zeta_{t,h}-\zeta_{t,h_{\max}}\right\|^2 = o\left(\Delta_n/\left(h_{\min}-\underline{h}\right)\right)+o(1)$ such that $|w_{73n}| = o_p(1)$ uniformly in $h\in H_n$.

To show (4.5) note that $\omega_{\hat{h}_n}-\omega_{h_{\max}} = l(\hat{h}_n)'\left(\Gamma_{\hat{h}_n}^{-1}\otimes\Sigma_v\right)l(\hat{h}_n) - l(h_{\max})'\left(\Gamma_{h_{\max}}^{-1}\otimes\Sigma_v\right)l(h_{\max})$ so that the same arguments used to show $E\left\|\zeta_{t,h}-\zeta_{t,h_{\max}}\right\|^2\to 0$ apply. For part i) of the theorem, note that $|\omega_h| \le \sum_{j_1,j_2=1}^{h}\left\|l'_{j_1}\right\|\left\|\left(\Gamma_h^{j_1,j_2}\otimes\Sigma_v\right)l_{j_2}\right\| \le \infty$ uniformly in $h$ such that it follows from absolute convergence arguments that $\omega_h\to\sum_{j_1,j_2=1}^{\infty} l'_{j_1}\left(\Gamma_h^{j_1,j_2}\otimes\Sigma_v\right)l_{j_2} < \infty$. The statement of part i) of the theorem then follows from applying the continuous mapping theorem to $\sqrt{n}w_{0n}(h_{\max})/\omega_{h_{\max}}$. ■

**Proof of Theorem (2.4):.** Let $c_n = (\log n/n)^{1/2}$. For all $\epsilon>0$, $P\left(\left\|\hat{\pi}(\hat{h}_n)-\pi(\hat{h}_n)\right\| > c_n\epsilon\right) \le P(\max_{h\in H_n}\|\hat{\pi}(h)-\pi(h)\| > c_n\epsilon)+o(1)$. Since $h\in H_n$ implies that $h\le o(\sqrt{n/\log n})$ it follows from An, Chen and Hannan (1982, p. 936) and Hannan and Kavalieris (1986, Theorem 2.1) that $\max_{h\in H_n}\sum_{j=1}^{h}\|\hat{\pi}_{j,h}-\pi_{j,h}\| \le \sum_{j=1}^{h_{\max}}\|\hat{\pi}_{j,h}-\pi_{j,h}\| = O_p((\log n/n)^{1/2})$. To see this note that as in the proof of Theorem 2.1 in Hannan and Kavalieris (1986, p.39), we have

$$\sum_{j=1}^{h_{\max}}\|\hat{\pi}_{j,h}-\pi_{j,h}\|\left\|\hat{\Gamma}_{j-k}^{yy}\right\| \le \sum_{j=0}^{h_{\max}}\|\pi_j\|\left\|\hat{\Gamma}_{j-k}^{yy}-\Gamma_{j-k}^{yy}\right\| + \sum_{j=1}^{h_{\max}}\|\pi_j-\pi_{j,h}\|\left\|\hat{\Gamma}_{j-k}^{yy}-\Gamma_{j-k}^{yy}\right\|$$

for $k=1,...,h_{\max}$ where $\sum_{j=1}^{h_{\max}}\|\pi_j-\pi_{j,h}\|\left\|\hat{\Gamma}_{j-k}^{yy}-\Gamma_{j-k}^{yy}\right\| = O_p(\sqrt{\log n/n})\sum_{h_{\max}}\|\pi_j\|$ by Hannan and Kavalieris (1986). Again, by Hannan and Deistler (1988, Theorem 7.4.3) it follows that

$$\sum_{j=0}^{h_{\max}}\|\pi_j\|\left\|\hat{\Gamma}_{j-k}^{yy}-\Gamma_{j-k}^{yy}\right\| = O_p(\sqrt{\log n/n}).$$

Since $\left\|\hat{\Gamma}_{j-k}^{yy}\right\| = O_p(1)$ uniformly by the same result it follows that $\sum_{j=1}^{h_{\max}}\|\hat{\pi}_{j,h}-\pi_{j,h}\| = O_p(\sqrt{\log n/n})$. Moreover, $\sum_{j=1}^{h}\|\pi_{j,h}-\pi_j\| = O(\sum_{j=h+1}^{\infty}\|\pi_j\|)$ by Hannan and Deistler (1988, Theorem 6.6.12). Since $h\ge h_{\min}$ and $h_{\min}$ satisfies $n^{1/2}\sum_{j=h_{\min}+1}^{\infty}\|\pi_j\|\to 0$ the result follows. ■

## References


Abadir, K. M., K. Hadri, and E. Tzavalis (1999): "The Influence of VAR Dimensions on Estimator Biases," *Econometrica*, 67, 163–181.

Akaike, H. (1969): "Power Spectrum Estimation Through Autoregressive Model Fitting," *Ann. Inst. Statist. Math.*, 21, 407–419.

An, H.-Z., Z.-G. Chen, and E. J. Hannan (1982): "Autocorrelation, Autoregression and Autoregressive Approximation," *Annals of Statistics*, 10, 926–936.

Andrews, D. W. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59(3), 817–858.

Bekker, P. A. (1994): "Alternative Approximations to the Distributions of Instrumental Variable Estimators," *Econometrica*, pp. 657–681.

Berk, K. N. (1974): "Consistent Autoregressive Spectral Estimates," *Annals of Statistics*, 2, 489–502.

Den Haan, W. J., and A. T. Levin (2000): "Robust Covariance Matrix Estimation with Data-Dependent VAR Prewhitening Order," *NBER Technical Working Paper No. 255*.

Donald, S. G., and W. K. Newey (2001): "Choosing the Number of Instruments," *Econometrica*, 69, 1161–1191.

Eastwood, B. J., and A. R. Gallant (1991): "Adaptive Rules for Seminonparametric Estimators that achieve Asymptotic Normality," *Econometric Theory*, 7, 307–340.

Goncalves, S., and L. Kilian (2003): "Asymptotic and Bootstrap Inference for AR(Infinity) Processes with Conditional Heteroskedasticity," *mimeo*.

Hahn, J., J. Hausman, and G. M. Kuersteiner (2000): "Bias Corrected Instrumental Variables Estimation for Dynamic Panel Models with Fixed Effects," *unpublished manuscript*.

Hahn, J., and G. Kuersteiner (2002): "Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects," *Econometrica*, 70, 1639–1657.

——— (2003): "Bias Reduction for Dynamic Nonlinear Panel Models with Fixed Effects," *mimeo*.

Hall, A. (1994): "Testing for a Unit Root in Time Series With Pretest Data-Based Model Selection," *Journal of Business & Economic Statistics*, 12, 461–470.

Hannan, E., and L. Kavalieris (1984): "Multivariate Linear Time Series Models," *Adv.Appl.Prob*, 16, 492–561.

——— (1986): "Regression, Autoregression Models," *Journal of Time Series Analysis*, 7, 27–49.

Inoue, A., and L. Kilian (2002): "Bootstrapping Smooth Functions of Slope Parameters and Innovation Variances in VAR(Infinity) Models," *International Economic Review*, 43, 309–331.

Kilian, L. (1998): "Small-Sample Confidence Intervals for Impulse Response Functions," *Review of Economics and Statistics*, 80, 218–230.

Kuersteiner, G. M. (2002): "Rate-Adaptive GMM Estimators for Linear Time Series Models," *manuscript.*

Leeb, H., and B. M. Pötscher (2003): "The Finite Sample Distribution of Post-Model-Selection Estimators and Unifrom Versus Nonuniform Approximations," *Econometric Theory*, 19, 100–142.

Lewis, R., and G. Reinsel (1985): "Prediction of Multivariate Time Series by Autoregressive Model Fitting," *Journal of Multivariate Analysis*, pp. 393–411.

Lütkepohl, H., and D. Poskitt (1996): "Testing for Causation Using Infinite Order Vector Autoregressive Processes," *Econometric Theory*, 12, 61–87.

Lütkepohl, H., and P. Saikkonen (1997): "Impulse Response Analysis in Infinite Order Cointegrated Vector Autoregressive Processes," *Journal of Econometrics*, 81, 127–157.

Ng, S., and P. Perron (1995): "Unit Root Tests in ARMA Models with Data-Dependent Methods for the Selection of the Truncation Lag," *Journal of the American Statistical Association*, 90, 268–281.

——— (2001): "Lag Length Selection and the Construction of Unit Root Tests With Good Size and Power," *Econometrica*, 69, 1519–1554.

Paparoditis, E. (1996): "Bootstrapping Autoregressive and Moving Average Parameter Estimates of Infinite Order Vector Autoregressive Processes," *Journal of Multivariate Analysis*, 57, 277–296.

Parzen, E. (1974): "Some Recent Advances In Time Series Modeling," *IEEE Trans. Automat. Control*, AC19, 723–730.

Pötscher, B. M. (1991): "Effects of Model Selection on Inference," *Econometric Theory*, 7, 163–185.

Saikkonen, P., and H. Lütkepohl (1996): "Infinite-order Cointegrated Vector Autoregressive processes: Estimation and Inference," *Econometric Theory*, 12, 814–844.

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

Sargan, J. (1975): "Asymptotic Theory and Large Models," *International Economic Review*, 16, 75–91.

Sen, P. K. (1979): "Asymptotic Properties of Maximum Likelihood Estimators Based on Conditional Specification," *Annals of Statistics*, 7, 1019–1033.

Shibata, R. (1980): "Asymptotically Efficient Selection of the Order of the Model for Estimating Parameters of a Linear Process," *Annals of Statistics*, 8, 147–164.

Shibata, R. (1981): "An Optimal Autoregressive Spectral Estimate," *Annals of Statistics*, pp. 300–306.