

The Influence of Covariance Hankel Matrix Dimension on Algorithms for VARMA Models

CELINA PESTANO-GABINO
CONCEPCIÓN GONZÁLEZ-CONCEPCIÓN
MARÍA CANDELARIA GIL-FARIÑA

Departamento de Economía Aplicada y Métodos Cuantitativos
Universidad de La Laguna (ULL)
38071 Campus de Guajara, La Laguna, SPAIN

Abstract: - Some methods for estimating VARMA models, and Multivariate Time Series Models in general, rely on the use of a Hankel matrix. Some authors suggest taking a larger dimension than theoretically necessary for this matrix. If the data sample is populous enough and the Hankel matrix dimension is unnecessarily large, this may result in an unnecessary number of computations, as well as in worse numerical and statistical results. We provide some theoretical results to know which is the Hankel matrix with the lowest dimension that is theoretically necessary and illustrate, with several simulated VARMA models, that using a dimension of the Hankel matrix greater than the theoretical minimal dimension proposed as valid does not necessarily lead to improved estimates. Although we use two algorithms, our main contributions are independent of the estimation method considered. We note that our paper does not include any comparisons between different algorithms for estimating VARMA models, as this is not our aim.

Key-Words: - Covariance Hankel matrices, Vector Autoregressive Moving-Average (VARMA) models, vector-valued linear stochastic systems, simulated models.

Received: April 17, 2019. Revised: December 9, 2019. Re-revised: January 15, 2020.
Accepted: January 22, 2020. Published: January 24, 2020.

1 Introduction

In general, the algorithms to estimate models for data analysis need to improve numerical and statistical efficiency. In particular, multivariate time series models, for example, vector autoregressive moving-average (VARMA) models, receive considerable attention by numerous authors. These models have a wide range of applications in industry, economy, biomedicine, physics... Many of the methods and techniques available are given, i.e., in [1] and in the reference therein. In this wide context, the efficiency condition presumably includes an important element, namely, reducing the dimension of the matrices involved in the calculations. Some methods for estimating the parameters of multivariate linear stochastic systems and, in general, for identifying and estimating multivariate models in time series analysis, involve a Hankel matrix of covariances (see, for instance, [1-12]). The rank of this matrix (the system order or McMillan degree) must be obtained and it is an important parameter for these different methods. In particular, it is used to set the dimension of certain matrices that intervene in various statistical algorithms proposed to estimate the models.

Our contribution concerns the influence of the choice of the Hankel matrix dimension on

identifying and estimating the model. This was driven by the observation that in the context of linear systems, VARMA models and so on, some authors proposed from the theoretical point of view unnecessary large for the dimensions of certain Hankel matrices, in particular, without being exhaustive:

In [2, p.216] we can read “the determination of the first basis of rows of the Hankel matrix $\rho_{s,s}$, where the integer s represents an upper bound on the dynamic dimension n of the process Z ”. In [5, p.2097] we can read, “Typically, one takes $m \approx (2$ to $4)n$. If no *a priori* guess of n is available, one can start with $m \approx N/10$ ” and in [6, Section 4.1.1], on System Identification (System ID), it states: “Note that in theory, we require the r used in determining the size of the Hankel matrix to be larger than the true order of the system. However, in practice, we often don’t know the true system order (...) Therefore, when we set the size of the Hankel matrix in our problems, as a rule of thumb, we use roughly twice the estimated order”. And, in Section 5, on Stochastic System Realization, “(...) as in the System ID problem, we need the Hankel matrix to be large enough, that is, j and k should be large than the rank” where j and k are the number of block-row and block-columns respectively. Also, in [6,

Appendix C] we can read “Assume further that r is larger than the true order of the system”. In [11, p.48] we can read “It may be useful to exploit the information contained in covariances corresponding to higher order lags in order to improve the quality of the estimator”.

Some theoretical results provided an insight into the properties of Hankel matrices associated with some models for multivariate time series and are useful to determine the bounds of theoretical/practical interest for the dimensions of these matrices.

By way of example, in this paper we have chosen two algorithms, in Section 2 the algorithm in [5] and in Section 4 the algorithm in [7], to illustrate, with simulated examples, the influence that the Hankel matrix dimension selected has on estimating VARMA models. Our conclusions are independent of the estimation method considered.

We do not intend to compare the algorithms, only to show that given a method that uses the aforementioned Hankel matrix, choosing a dimension greater than the minimum required does not always lead to better estimates for the parameters of the model in question. Therefore the computational work could be substantially reduced by using the considerations proposed in this paper.

As in other similar contexts, it is very important to have a reasonably reliable algorithm to calculate the rank of the matrix and to stabilize the matrices involved in the process.

This paper contains four sections. In Section 2, we present the use of a Hankel matrix in one of the algorithms for estimating VARMA models ([5]). In Section 3 we propose a set of criteria and indicators that allow us to illustrate that if the sample size is large enough and the dimension chosen for the Hankel matrix is unnecessarily large, this may result in a superfluous number of computations as well as in numerical problems. In Section 4, we present Extended Yule Walker (XYW) method so as to illustrate with another algorithm (see [7]) the influence of the Hankel matrix dimension in practice. We finalize with the conclusions, acknowledgements and references.

2 Some Considerations on the Algorithm in [5]

In [5] the authors work with a discrete-time square-integrable stationary, vector-valued stochastic process $y(t)$, for t integer. They assume that this process has a rational spectral density although it will be apparent in the sequel that the identification method will return valid finite-

dimensional models even when this nontestable assumption fails. Under these conditions, the process admits a finite-dimensional state-space innovation representation as follows:

$$\begin{cases} u(t+1) = Au(t) + Ke(t) & (n \times 1) \\ y(t) = Cu(t) + e(t) & (d \times 1) \end{cases} \quad (1)$$

Here $u(t)$ is the state vector, and A, K, C are matrices of appropriate dimensions, whereas the vector-valued sequence $e(t)$ is the white innovation sequence. They assume zero-mean processes throughout. The innovation covariance matrix is denoted by Q

$$E[e(t)e^T(s)] = Q \delta_{t,s}$$

where E denotes the expected value with respect to the underlying probability measure, and $\delta_{t,s}$ is the Kronecker delta function. It is assumed that Q is strictly positive definite and that (1) is a minimum degree representation of $y(t)$ in the stochastic sense, i.e., n is the McMillan degree.

It is well-known ([1]) that, for certain non negative integers p and q , the process $y(t)$ could be written as a VARMA(p, q) representation

$$y(t) + A_1 y(t-1) + \dots + A_p y(t-p) = e(t) + B_1 e(t-1) + \dots + B_q e(t-q)$$

where A_i and B_i are $d \times d$ matrices and, denoting $A_p(z) = I + A_1 z + \dots + A_p z^p$ and $B_q(z) = I + B_1 z + \dots + B_q z^q$, the VARMA transfer matrix could be written as $F(z) = A_p^{-1}(z) B_q(z)$ or, in equivalent way, $F(z) = C(zI - A)^{-1} K + I$ if we consider (1).

In [5] a VARMA parameter estimation method is introduced. The problem is to estimate the order n and the matrices A, K, C and Q , from N observations of a sample realization of $y(t)$, in such a way that the estimated model matches the second order properties of the process $y(t)$.

The process covariances in [5] are denoted by $R_k = E[y(t)y^T(t-k)]$ for t, k integers, and the Hankel matrix of covariances by R , that is [5, p.2095]:

$$R = \begin{pmatrix} R_1 & R_2 & \dots & R_m \\ R_2 & R_3 & \dots & R_{m+1} \\ \dots & \dots & \dots & \dots \\ R_m & R_{m+1} & \dots & R_{2m-1} \end{pmatrix}$$

The method in [5] involves a series of steps. We focus on a part of the first step. Specifically, the first step is done through a Singular Value

Decomposition (SVD) of the Hankel matrix R which is a resource common to an important vein of so-called subspace methods (see, for instance, references in [5]).

The authors consider the SVD, $R = U\Sigma V^T$, where Σ is a $n \times n$ nonsingular diagonal matrix and $UU^T = V^T V = I_n$, the *observability* and the *controllability matrices*, $\Omega = U\Sigma^{1/2}$ and $\Gamma^T = \Sigma^{1/2} V^T$, respectively. In particular, the matrix A is obtained as the solution to the following system:

$$(\overline{U}\Sigma^{1/2})A = \underline{U}\Sigma^{1/2} \quad (2)$$

where \overline{U} and \underline{U} are the matrices made from the first and the last $m-1$ block rows of U , respectively.

2.1 Theoretical Considerations on Matrix Hankel Dimension in the Algorithm in [5]

Note that the Hankel matrix R depends on m . Therefore, in our paper we prefer to denote this matrix by $R(m)$.

Remark: For our aim, we denote $s^* = \{\min m / \text{rank } R(m) = n\}$ and remark that: s^* does not represent an upper bound on the dynamic dimension n of the process because for some VARMA process $s^* < n$. We can observe it in the following example:

Example 1. We consider, $Q = \begin{pmatrix} 4 & 1 \\ 1 & 1 \end{pmatrix}$ and $F(z) = \sum_{i=0}^{\infty} F_i z^i$ where $F_0 = I$, $F_1 = -I$, $F_2 = \begin{pmatrix} 6 & 2 \\ 2 & 0 \end{pmatrix}$ and $F_i = 0.5F_{i-2}$ if $i \geq 3$. The autocovariance matrices of the VARMA(2,2) process $y(t)$ are:

$$R_0 = I, R_{2i+1} = \begin{pmatrix} 0.5^i & 0 \\ 0 & 0.5^i \end{pmatrix}, R_{2i+2} = 0 \text{ if } i \geq 0 .$$

In this example $n=4$, $s^*=2$, therefore $s^* < n$, and s^* does not represent an upper bound on the dynamic dimension.

In this context, the possibility $s^* \leq n$ could have some consequences on the theory-practice duality. For instance, the theoretical condition $m \geq n$ included in [5, p.2094-2095] could be improved to $m \geq s^* + 1$.

With the following propositions we propose the minimum value of m such that (2) has a unique solution for A i.e., the algorithm in [5] works.

Proposition 1: Considering R with $m = s^* + 1$, i.e., $R(s^* + 1)$ in the first step of the procedure in [5], we can ensure that the *observability* and the *controllability matrices* are both full rank, n .

Proof. In the first step of the procedure in [5], to estimate the linear system parameter, a Singular Value Decomposition (SVD) of R is obtained. Given any matrix G , if we obtain a SVD, $G = XSY^T$, it is well known that X is full (column) rank and Y^T is full (row) rank.

Let us denote a SVD of $R(m) = U(m)\Sigma(m)V^T(m)$ and, in a similar way of [5], $\Omega(m) = U(m)\Sigma^{1/2}(m)$ and $\Gamma^T(m) = \Sigma^{1/2}(m)V^T(m)$. Therefore, $\forall m \geq 1$,

$$\text{rank } R(m) = \text{rank } \Omega(m) = \text{rank } \Gamma^T(m) \quad (3)$$

Note that, $\text{rank } R(m) < n$ if $m < s^*$, and $\text{rank } R(m) = n$ if $m \geq s^*$. From (3), $\text{rank } \Omega(m) = \text{rank } \Gamma^T(m) = n$, $\forall m \geq s^*$. Since $s^* \leq n$, then $\text{rank } \Omega(m) = \text{rank } \Gamma^T(m) = n$, $\forall m \geq n$. Taking

into account that in [5], $\Omega(s^*) = \begin{pmatrix} C \\ CA \\ \vdots \\ CA^{s^*-1} \end{pmatrix}$ and

$\Gamma^T(s^*) = (K \dots AK^{s^*-1})$, the condition $m \geq s^*$ ensures that the extended observability matrix $\Omega(m)$ has full (column) rank and that the extended controllability matrix $\Gamma^T(m)$ has full (row) rank. \square

Proposition 2. Considering the SVD of $R(s^* + 1)$, the system (2) has a unique solution for A .

Proof. Note that $R(s^* + 1)$ is the Hankel matrix with the lowest dimension that is theoretically necessary in the procedure in [5]. It is because $\text{rank } R(s^*) = \text{rank } R(m)$ for any $m \geq s^*$. Therefore in [5, p.2095], the system $(\overline{U}\Sigma^{1/2})A = \underline{U}\Sigma^{1/2}$ has a single solution, if \overline{U} and \underline{U} are the matrices made from the first and last s^* block rows of U , respectively. \square

Remarks:

- In theory, all the series of real covariances are necessary to calculate n , and, as a consequence, in practice a large number of estimated covariances is necessary to obtain a suitable estimate of n . However, if s^* is known, only the $2s^* + 1$ covariances included in $R(s^* + 1)$ are sufficient to calculate (A, K, C, Q) with the theoretical procedure in [5].
- Note that Propositions 1 and 2 above could be used to improve the algorithm in [5]. Of course, it does not contradict it. For instance, in Example 1, A can be estimated from the Hankel matrix $R(3)$ and $3 < n$.
- Throughout our paper we do not question that $\text{rank } R(m) = n$, $\forall m \geq n$. This is true. What we

affirm is that there exists $s^* \leq n$ such that rank $R(m) = n \quad \forall m \geq s^*$.

2.2 Practical Considerations on Hankel Matrix Dimension in the Algorithm in [5]

To illustrate that given a method that uses the aforementioned Hankel matrix, choosing a dimension greater than the minimum required does not always lead to better estimates for the parameters of the model in question, several VARMA models were computed. Since all of the models showed that our theoretical-practical proposal is valid, we thought it unnecessary to provide detailed results of every example, instead selecting three that we regard as representative of what we are proposing.

In every example we simulated, $e(t)$ is a vector white-noise process with covariance matrix

$$Q = \begin{pmatrix} 4 & 1 \\ 1 & 1 \end{pmatrix} \text{ and we used a file with the simulated set}$$

$\{e(1), \dots, e(35000)\}$. The number of executions for each example was $S = \lceil 35000/T \rceil$, where T denotes the sample size chosen for $e(t)$ in each execution. For instance, if $T = 350$, we have 100 executions for each example. Moreover, for each execution, we ignored, as is customary, the first values of the process to have $N = T - 10$. We consider several sample sizes (N).

Now, we illustrate that if \hat{n} is a suitable value of n , estimated using some previous statistical procedure, and $\hat{s}^* = \min\{m / \text{rank } \hat{R}(m) = \hat{n}\}$, then the $2\hat{s}^* + 1$ estimated covariances in $\hat{R}(\hat{s}^* + 1)$ are sufficient to estimate (A, K, C, Q) with the practical procedure in [5]. However, if in practice there is uncertainty as to the value of \hat{s}^* , then we propose considering $\hat{R}(\hat{n} + 1)$ to estimate (A, K, C, Q) .

The MATLAB algorithm mentioned in [5] considers $R(m^*)$, where $m^* = \lceil u/2 \rceil$, $u = \min\{6n, \lceil 3(N/\log_{10}N)^{1/2} \rceil\}$, n the order of the model and N the sample size.

We will use several simulated examples to illustrate that our proposal to use the Hankel matrix with the lowest dimension that is theoretically required, namely $\hat{R}(\hat{s}^* + 1)$, remains valid for at least certain criteria and indicators. We will use the algorithm proposed in [5], changing only the dimension of the Hankel matrix used but considering several dimensions, from $R(s^* + 1)$ to $R(m^*)$, that is, from the one that we propose up to the default value m^* in the MATLAB procedure in [5]. As we will see, there are no substantial improvements as a consequence of using Hankel

matrices with a higher dimension than the one we propose.

3 Criteria and Indicators based on the Innovation Covariance Matrix Q

To assess the above proposal from a practical standpoint, we will use three criteria based on certain errors in estimating the innovation covariance matrix Q , the white-noise autocovariance matrix. These criteria allow us to compare the estimates obtained for various sizes of the Hankel matrix present in the estimation process, and determine whether any improvements result from increasing the size of this matrix beyond the bound proposed in the theory.

Given a set of data $\{y(1), \dots, y(M)\}$, to estimate the matrix $Q = (Q_{ij})_{i,j=1,2}$ as per [5], it is necessary to first estimate the matrices A , C and K [5, p.2095]. In this section, the data come from simulated models. Once the model is chosen, assume we carry out S executions of the model and let $W = \{1, 2, \dots, S\}$.

For each value m , from $s^* + 1$ to m^* , we will:

- Estimate Q for each execution, using the algorithm proposed in [5] and taking Hankel matrix $R(m)$. This estimate is denoted by $Q(m, x)$, for $x = 1, \dots, S$.
- Calculate $G(m) = \{x \in W / \|Q - Q(m, x)\| \leq 90 \text{ and } \|Q - Q(m^*, x)\| \leq 90\}$ where $\|\cdot\|$ denotes the norm 2 of a matrix. We arbitrarily chose 90 because, with $G(m)$ our goal is to select those executions that do not have overly high values, whether we consider $R(m)$ or $R(m^*)$ in the estimates, without quite eliminating all those executions that we consider to be "not suitable".
- Calculate $\#G(m)$ as the cardinal of $G(m)$ set.

Next, based on these results, we will assess the three criteria in Table 1 based on Q , which measure the mean error in matrix Q , the mean squared error of each element in said matrix and what we define as outliers.

Table 1. Criteria based on Q

	Criteria
1 (Mean error in Q)	$EM(m) = \frac{\sum_{x \in G(m)} \ Q - Q(m, x)\ }{\#G(m)}$
2 (Mean squared error in element Q_{ij})	$ECMQ_{ij}(m) = \frac{\sum_{x \in G(m)} (Q_{ij} - Q_{ij}(m, x))^2}{\#G(m)}$ for $i, j = 1, 2$
3 (Outliers)	$Outlier(m) = \#\{x \in W / \ Q - Q(m, x)\ > 1\}$, that is, the number of simulations that yielded $\ Q - Q(m, x)\ > 1$

Note that we eliminated from the formulas $EM(m)$ and $ECMQ_{ij}(m)$ those executions in which abnormally large estimates appear. We used $\#G(m)$ to denote the actual number of executions used to calculate the mean errors.

We used these criteria to attempt to identify the goodness of the estimate as a function of the value of m by using four indicators. Specifically, we assessed the best comparisons that are obtained for two different values ($m=a$ and $m=b$) once the S estimates of Q are completed assuming $m=a$, and the S estimates of Q assuming $m=b$.

Indicator 1: The value $m=a$ is better than $m=b$ in terms of the mean error iff $EM(a) < EM(b)$.

Indicator 2: The value $m=a$ is better than $m=b$ in terms of the mean squared error of Q_{ij} ($i, j=1, 2$) iff $ECMQ_{ij}(a) < ECMQ_{ij}(b)$.

Indicator 3: The value $m=a$ is better than $m=b$ in terms of the number of outliers iff $Outlier(a) < Outlier(b)$.

Indicator 4: Percentage of the S executions in which $m=a$ is better than $m=b$, in the sense that, for a particular execution x , $\|Q-Q(a,x)\| < \|Q-Q(b,x)\|$ iff $m=a$ is better than $m=b$ for said execution. Similarly, Indicator 4 is considered for each Q_{ij} .

4 Simulated Example Results

The simulated examples we will discuss in this section are in Table 2.

Table 2. Simulated Examples

Example	Model	n	s^*
2	$y(t) = e(t) + \begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix} e(t-2)$	2	2
3	$y(t) - \begin{pmatrix} 1/1.2 & 0.5 \\ 0 & 1/1.2 \end{pmatrix} y(t-2) = e(t)$	4	2
4	$y(t) + \begin{pmatrix} 0 & 0 \\ 0 & 1/2 \end{pmatrix} y(t-1) = e(t) + \begin{pmatrix} 0 & 0 \\ 0 & 1/2 \end{pmatrix} e(t-1) + \begin{pmatrix} 1/8 & 1/4 \\ 0 & 0 \end{pmatrix} e(t-2)$	3	2

As for the sample sizes, we started out considering $T=100$ ($N=90$), 200 ($N=190$), 300 ($N=290$) and 400 ($N=390$). As for the order, we assumed a possible error by letting the estimated order be either $n-1$ or $n+1$ instead of n , which is its true value.

We noticed that a sample size of $N=90$ was not suitable, since all of the simulated examples yielded a high value for $Outlier(m)$ (Criterion 3). As the sample size was increased, this value of $Outlier(m)$ decreased, as expected, and all of the indicators improved. Specifically, and by way of example, Tables 3-7 associated with Example 2 are presented:

Table 3. Example 2. $Outlier(m)$

	T	N	$m=3$	$m=4$	$m=5$	$m=6$
350	100	90	111	111	110	116
175	200	190	5	5	5	7
116	300	290	2	1	1	1
87	400	390	0	0	0	0

Table 4. Example 2. $EM(m)$

S	T	N	$m=3$	$m=4$	$m=5$	$m=6$
350	100	90	0.8444	0.8396	0.8384	0.8505
175	200	190	0.4959	0.5030	0.5058	0.5081
116	300	290	0.3686	0.3705	0.3718	0.3688
87	400	390	0.2985	0.3066	0.3052	0.3056

Table 5. Example 2. $ECMQ_{11}(m)$

S	T	N	$m=3$	$m=4$	$m=5$	$m=6$
350	100	90	0.4849	0.4898	0.4794	0.4860
175	200	190	0.2051	0.2107	0.2093	0.2146
116	300	290	0.1115	0.1140	0.1120	0.1116
87	400	390	0.0706	0.0755	0.0738	0.0740

Table 6. Example 2. $ECMQ_{12}(m)$

S	T	N	$m=3$	$m=4$	$m=5$	$m=6$
350	100	90	0.1662	0.1629	0.1657	0.1691
175	200	190	0.0538	0.0531	0.0550	0.0544
116	300	290	0.0318	0.0307	0.0315	0.0309
87	400	390	0.0214	0.0223	0.0227	0.0222

Table 7. Example 2. $ECMQ_{22}(m)$

S	T	N	$m=3$	$m=4$	$m=5$	$m=6$
350	100	90	0.1900	0.1842	0.1832	0.2021
175	200	190	0.0418	0.0440	0.0417	0.0405
116	300	290	0.0187	0.0184	0.0194	0.0197
87	400	390	0.0102	0.0103	0.0100	0.0096

We can state that not only in Example 2, but in every case we studied, for any of the sample sizes used and the indicators considered, increasing the size of the Hankel matrix yielded no substantial improvement, meaning there were no reductions in any of the criteria used (errors and outliers).

In light of the results and since the noise files we simulated have 35000 bivariate pieces of data, the sample size used from here on is $N=340$ ($T=350$), which we believe to be an adequate sample size.

Having set the sample size, let us consider what happens if we start with an erroneous value for the order of the system, n . If r denotes the estimated value of n , the results for Example 2 are in Table 8, where the true order is $n=2$ and the minimum theoretical size of the Hankel matrix is $s^*+1=3$.

Table 8. Example 2. *Outlier(m)*

	$r=1$ $m^*=3$	$r=n=2$ $m^*=6$	$r=3$ $m^*=9$	$r=4$ $m^*=12$
$m=r$	100	81	5	12
$m=r+1$	100	0	2	12
$m=r+2$	100	0	1	18
$m=r+3$		0	4	12
$m=r+4$		0	5	11
$m=r+5$			1	15
$m=r+6$			7	12
$m=r+7$				14
$m=r+8$				15

In this example, the best results based on Criterion 3 (smallest number of outliers) were obtained for values close to, but not below, the actual value of n and the minimum theoretical value of m . We obtained too many outliers when we used a value for n that was below the true value (100 outliers), or when we selected a proper n but used a Hankel matrix size that was below the minimum theoretical value (81 outliers). When we used r larger than n , we obtained fewer outliers if $r=n+1$ (between 5 and 7). The number of outliers increased slightly if $r=n+2$ was used (between 12 and 18). There were no outliers when we used the correct value of n and a theoretically valid value of m .

While the outcome was not exactly the same in every example, we can state that in all of them, setting r close to the true value n and a matrix size higher than the rank value did not yield an improvement in terms of Criterion 3 (lower number of outliers) after the size of the Hankel matrix was increased beyond its theoretically valid value.

When interpreting the following indicators, it makes no sense to consider the cases with many outliers in the previous table (the ones in the first column, $r=1$, and in the box $r=2$ and $m=r$), which are located in positions that precede their theoretically proposed positions.

The results allows us to state that, for a suitable sample size, even if the estimated value of n is erroneous, but close to the real value, using a Hankel matrix of a dimension larger than the smallest theoretically valid size does not improve any of the indicators analyzed.

Table 9. Example 2. *EM(m)*

	$r=1$ $m^*=3$	$r=n=2$ $m^*=6$	$r=3$ $m^*=9$	$r=4$ $m^*=12$
$m=r$	13.8108	7.6535	4.0162	1.1586
$m=r+1$	13.6165	0.3630	0.5293	1.8786
$m=r+2$	8.1807	0.3660	0.3941	1.7626
$m=r+3$		0.3711	1.0377	0.9093
$m=r+4$		0.3675	30.4109	1.6698
$m=r+5$			0.3943	0.6395
$m=r+6$			0.8124	14.1407
$m=r+7$				1.0448
$m=r+8$				1.2209

Table 10. Example 2. *ECMQ₁₁(m)*

	$r=1$ $m^*=3$	$r=n=2$ $m^*=6$	$r=3$ $m^*=9$	$r=4$ $m^*=12$
$m=r$	47.4094	23.6179	0.3174	0.2907
$m=r+1$	44.6208	0.1067	0.1408	0.4154
$m=r+2$	16.3536	0.1065	0.1128	0.5334
$m=r+3$		0.1084	0.1472	0.3189
$m=r+4$		0.1065	0.3183	0.5210
$m=r+5$			0.1066	0.3929
$m=r+6$			0.1544	0.5478
$m=r+7$				0.3934
$m=r+8$				0.4398

Table 11. Example 2. *ECMQ₁₂(m)*

	$r=1$ $m^*=3$	$r=n=2$ $m^*=6$	$r=3$ $m^*=9$	$r=4$ $m^*=12$
$m=r$	48.2302	22.4074	0.2873	0.2305
$m=r+1$	47.6787	0.0307	0.0474	0.6718
$m=r+2$	17.2733	0.0313	0.0350	2.7610
$m=r+3$		0.0331	0.1457	0.4431
$m=r+4$		0.0329	0.1542	2.8267
$m=r+5$			0.0379	0.2152
$m=r+6$			0.1654	3.3151
$m=r+7$				0.3602
$m=r+8$				0.6236

Table 12. Example 2. $ECMQ_{22}(m)$

	$r=1$ $m^*=3$	$r=n=2$ $m^*=6$	$r=3$ $m^*=9$	$r=4$ $m^*=12$
$m=r$	48.6383	22.4391	1227.7899	27.1625
$m=r+1$	50.6689	0.0135	1.1742	116.9012
$m=r+2$	18.1347	0.0141	0.0369	101.4346
$m=r+3$		0.0137	38.8910	6.9426
$m=r+4$		0.0145	89991.183	52.9457
$m=r+5$			0.0530	0.1806
$m=r+6$			7.8917	12.760.525
$m=r+7$				13.284
$m=r+8$				37.428

In Tables 9-12, it is important to keep in mind that the large values in some boxes are due to the outliers associated with them.

In Table 12 we see that for some cases for which the estimated rank is incorrect, the high number of outliers significantly distorts $ECMQ_{22}$. If we omit those simulations of x such that $\|Q-Q(m,x)\|>1$ or $\|Q-Q(m^*,x)\|>1$, we obtain the Table 13, in which we have ignored the boxes with a large number of outliers.

Table 13. Example 2. $ECMQ_{22}(m)$ ignoring outlier executions

	$r=n=2$ $m^*=6$	$r=3$ $m^*=9$	$r=4$ $m^*=12$
$m=r$		0.0202	0.0441
$m=r+1$	0.0135	0.0275	0.0385
$m=r+2$	0.0141	0.0225	0.0210
$m=r+3$	0.0137	0.0288	0.0282
$m=r+4$	0.0145	0.0335	0.0291
$m=r+5$		0.0343	0.0164
$m=r+6$		0.0257	0.0247
$m=r+7$			0.0273
$m=r+8$			0.0345

Tables 14-21 show a summary of the results from the two other examples chosen to illustrate our proposal.

Table 14. Example 3. $n=4, s^*+1=3, m^*=12$

	Outlier (m)	EM (m)	$ECMQ_{11}$ (m)	$ECMQ_{12}$ (m)	$ECMQ_{22}$ (m)
$m=3$	1	0.3714	0.1429	0.0188	0.0079
$m=4$	1	0.3721	0.1429	0.0190	0.0080
$m=5$	1	0.3715	0.1412	0.0186	0.0081
$m=6$	1	0.3738	0.1422	0.0193	0.0086
$m=7$	1	0.3701	0.1391	0.0190	0.0086
$m=8$	1	0.3726	0.1390	0.0192	0.0091
$m=9$	1	0.3721	0.1409	0.0194	0.0095
$m=10$	1	0.3696	0.1400	0.0181	0.0094
$m=11$	1	0.3745	0.1437	0.0186	0.0092
$m=12$	1	0.3756	0.1475	0.0190	0.0094

Table 15. Example 3¹. $Outlier(m)$

	$r=3$	$r=4$	$r=5$
$m=5$	9	1	8
$m=6$	9	1	7
$m=7$	8	1	10
$m=8$	9	1	6

Table 16. Example 3. $EM(m)$

	$r=3$	$r=4$	$r=5$
$m=5$	0.6502	0.3802	0.7038
$m=6$	0.6371	0.3826	40.7745
$m=7$	0.6389	0.3786	24.5716
$m=8$	0.6404	0.3813	10.6669

Table 17. Example 3. $ECMQ_{11}(m)$

	$r=3$	$r=4$	$r=5$
$m=5$	0.2011	0.1522	0.2289
$m=6$	0.2095	0.1536	0.2090
$m=7$	0.2052	0.1501	0.2021
$m=8$	0.2164	0.1500	0.1892

Table 18. Example 3. $ECMQ_{12}(m)$

	$r=3$	$r=4$	$r=5$
$m=5$	0.0571	0.0198	0.0438
$m=6$	0.0538	0.0205	0.1021
$m=7$	0.0542	0.0200	0.0506
$m=8$	0.0520	0.0205	0.1507

¹ In some of the simulations the MATLAB algorithm was interrupted due to, among other reasons, improper conditioning resulting from using a small sample size or an incorrect rank. The values shown in the table were calculated without interrupting the execution.

Table 19. Example 3. $ECMQ_{22}(m)$

	$r=3$	$r=4$	$r=5$
$m=5$	0.2485	0.0082	4.4144
$m=6$	0.2328	0.0088	148149.7
$m=7$	0.2428	0.0087	37353.5
$m=8$	0.2406	0.0092	9165.7

The abnormally high values in Table 19 are associated with an erroneous rank value.

Table 20. Example 4. $n=3, s^*+1=3, m^*=9, T=350$

m	Outlier (m)	EM (m)	$ECMQ_{11}$ (m)	$ECMQ_{12}$ (m)	$ECMQ_{22}$ (m)
3	8	662.0904	0.1503	0.1544	43710823.5
4	2	0.3532	0.1327	0.0176	0.0071
5	1	0.3378	0.1195	0.01597	0.0063
6	1	0.3415	0.1251	0.0163	0.0068
7	2	0.3430	0.1236	0.0166	0.0082
8	3	87.8	0.1218	0.0506	765116.1
9	2	0.3438	0.1238	0.0175	0.0077

Regarding Table 20 and its high values, if the sample size is increased to 500, only one outlier results, the simulation of which we omitted from the calculation of the criteria to yield the results in Table 21.

Table 21. Example 4. $n=3, s^*+1=3, m^*=9, T=500$

m	Outlier (m)	EM (m)	$ECMQ_{11}$ (m)	$ECMQ_{12}$ (m)	$ECMQ_{22}$ (m)
3	1	0.3087	0.0886	0.0122	0.0162
4	0	0.2831	0.0783	0.0123	0.0050
5	0	0.2783	0.0779	0.0119	0.0041
6	0	0.2777	0.0786	0.0116	0.0040
7	0	0.2756	0.0765	0.0115	0.0039
8	0	0.2754	0.0755	0.0118	0.0042
9	0	0.2728	0.0755	0.0118	0.0040

We thus deduced that for this example, a sample size of 500 is better suited than the 350 we selected for the bulk of our experiment, though 350 was adequate for almost all of the examples analyzed.

To conclude, we will evaluate Indicator 4 using $m=s^*+1$ and $m=s^*+2$ in order to compare each of the estimates $Q(s^*+1, x)$ and $Q(s^*+2, x)$, with $Q(m^*, x)$ for $x \in W$.

The Table 22 contains the values obtained for the three examples chosen, using $T=350$. The interpretation of the table is as follows. In $p\%$ of the

S executions, Q was estimated better using $m=s^*+1$ instead of $m=m^*$, if $\|Q-Q(s^*+1, x)\| < \|Q-Q(m^*, x)\|$ for the $p\%$ of values of x . Similarly, in $p\%$ of the S executions, Q_{ij} was better using $m=s^*+1$ instead of $m=m^*$, if $|Q_{ij}-Q_{ij}(s^*+1, x)| < |Q_{ij}-Q_{ij}(m^*, x)|$ for the $p\%$ of values of x .

Table 22. Results of Indicator 4. Using $T=350$

	Q	Q_{11}	Q_{12}	Q_{22}
Example 2, $m=s^*+1$	49%	53%	51%	59%
Example 2, $m=s^*+2$	44%	45%	56%	59%
Example 3, $m=s^*+1$	56%	57%	55%	61%
Example 3, $m=s^*+2$	54%	57%	50%	61%
Example 4, $m=s^*+1$	42%	48%	47%	42%
Example 4, $m=s^*+2$	44%	48%	47%	52%

Therefore, for Example 2, the estimate for Q_{11} was better using $m=s^*+1$ instead of $m=m^*$ in 53% of the executions; in other words,

$|Q_{11}-Q_{11}(s^*+1, x)| < |Q_{11}-Q_{11}(m^*, x)|$ for 53% of the executions. In 45% of the cases $m=s^*+2$ was better than $m=m^*$; that is,

$|Q_{11}-Q_{11}(s^*+2, x)| < |Q_{11}-Q_{11}(m^*, x)|$ in 45% of the executions. In Example 3, the estimate of Q was better using $m=s^*+1$ instead of $m=m^*$ in 56% of the executions, and so on.

In all of the examples analyzed, we find that these values hover around 50%, which confirms, for these cases, our hypothesis that the indicators considered do not improve as the size of the Hankel matrix is increased.

All of the examples also show that when the sample size is increased, the percentages tend to stabilize even more, if possible, around 50%. This means that given a sufficiently large sample size, the same estimate is obtained with $R(s^*+1)$ as with $R(m^*)$, but with considerably less computational work, and thus greater efficiency.

Briefly, and so as to illustrate our contributions with another algorithm, in the following section we consider XYW method.

4 Some Considerations on XYW Method

XYW equations have been proposed in [9], and recently considered, i.e., in [7,8,10,11] for estimation of VAR and VARMA parameters. Under certain conditions, the parameters of a VAR or VARMA model are determined uniquely with available population covariance of single –or mixed–

frequency observations on the variables of the model.

To study mixed-frequency data (MFD) cases, it is usual (see for instance [10]) to consider a VARMA model of $d=d_1+d_2$ variables, whose first d_1 variables are high-frequency variables observed in every period and whose last d_2 variables are low-frequency variables observed every certain number of periods. Many MFD cases can be handled as in this MFD case. For mixed-frequency data in a more general form, see for instance [11].

For our illustrative purpose, we have chosen the case of stationary and regular VARMA($p,0$) models denoted by

$$y(t)=A_1y(t-1)+\dots+A_p y(t-p)+e(t)$$

and the innovation variance is denoted by Q . Let \tilde{C}_i the first d_1 columns of R_i .

Denoting the Hankel matrix

$$D_L = \begin{pmatrix} \tilde{C}_0^T & \dots & \tilde{C}_{p-1}^T \\ \vdots & & \vdots \\ \tilde{C}_{L+p-1}^T & \dots & \tilde{C}_{L+2p-2}^T \end{pmatrix}, \quad X=-(A_p^T \dots A_1^T)^T \text{ and}$$

$E=(\tilde{C}_p^T \dots \tilde{C}_{L+2p-1}^T)^T$, if $\exists L/\text{rank } D_L=dp$, we can consider the XYW estimator

$$X=(D_L^T D_L)^{-1} D_L^T E_L \quad (4)$$

and $\text{vec}(Q)=$

$$((G \otimes G)(I_{(dp)^2} - (A \otimes A))^{-1} (G^T \otimes G^T))^{-1} \text{vec } C_0$$

with $G=(I_d, 0, \dots, 0)$, $A = \begin{pmatrix} A_1 & \dots & A_{p-1} & A_p \\ I_d & & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & I_d & 0 \end{pmatrix}$. See,

for instance, [7].

4.1 The Influence of the Hankel Matrix Dimension in Practice: XYW Method

We now consider the influence of the Hankel matrix dimension on the estimate for the parameters of a VARMA(1,0) model, using the example below.

Example 5:

$$y(t)-A_1y(t-1)=e(t) \text{ with } A_1 = \begin{pmatrix} 0.9556 & 0.8611 \\ -0.6914 & 0.2174 \end{pmatrix}.$$

We use the simulated set $\{e(1), \dots, e(35000)\}$ in Section 2, consider two cases ($d_1=2$ and $d_1=1$) and compare the results obtained for different values of L with those corresponding to a ‘‘sufficiently large’’ arbitrary value, chosen for illustrative purposes only, such as 11.

Since the XYW estimators tend to be used as the initial estimators for other estimating methods (see for instance [11]), we will use the mean error in A for each value of $L=1, \dots, 11$

$$EMA_1(L) = \frac{\sum_{x \in W} \|A_1 - A_1(L, x)\|}{\#W}$$

where $A_1(L, x)$ is the XYW estimate of A_1 for execution x , considering Hankel matrix D_L in (4).

For Example 5, we find that $\{x \in W / \text{if } |1-\lambda A_1(L, x)|=0 \text{ then } |\lambda|>1\}=W$, therefore all the estimated A_1 are stable.

Case (I): $d_1=d=2$ (i.e., the same frequencies are present for both variables)

In D_L we let $\sum_{t=1}^N y(t)y^T(t-i)$ instead of \tilde{C}_i for $i=0, \dots, L+2p-1$.

We include the results in Table 23. Note that the best results are always obtained with the lowest value of L , and that essentially, the results are not affected if we select another value ‘‘sufficiently large’’ for L .

Table 23. Case (I) $d_1=d=2$

	$EMA_1(L)$ $T=350$	$EMA_1(L)$ $T=700$	$EMA_1(L)$ $T=1000$
$L=1$	0.04391325	0.03232048	0.02512431
$L=2$	0.04453883	0.03280759	0.02477276
$L=3$	0.04502489	0.0330484	0.02466068
$L=4$	0.04543278	0.03334947	0.02470456
$L=5$	0.04632424	0.03414898	0.0252193
$L=7$	0.04901777	0.03598402	0.02674568
$L=11$	0.05168144	0.03835419	0.02920605

Case (II): $d_1=1, d=2$ (assuming we have all the frequencies of the first variable and, for example, only the odd frequencies of the second variable).

In D_L let $\begin{pmatrix} \sum_{t=1}^N y_1(t)y_1^T(t-i) \\ \sum_{t=1}^{(N+1)/2} y_2(2t-1)y_1^T(2t-1-i) \end{pmatrix}$ instead of

\tilde{C}_i for $i=0, \dots, L+2p-1$ (see, for instance, [9]).

We obtain the results in Table 24.

Table 24. Case (II) $d_1=1$

	$EMA_1(L)$ $T=350$	$EMA_1(L)$ $T=700$	$EMA_1(L)$ $T=1000$
$L=1$	0.08390932	0.06064302	0.0522413
$L=2$	0.0569018	0.04406534	0.03614363
$L=3$	0.05064516	0.03783381	0.02973002
$L=4$	0.05129033	0.03802203	0.03012609
$L=5$	0.05276242	0.03827057	0.03069578
$L=7$	0.05436977	0.04017285	0.03164954
$L=11$	0.0570004	0.04193298	0.03353029

Reading Table 24 by columns we observe that for sample sizes of 350, 700 and 1000, the best results for $EMA(L)$ are for $L=3$.

As expected, if we fix the value for L , i.e., reading Table 24 by columns, increasing the sample size results in $EMA_i(L)$ decreasing.

Therefore, our thesis, that expanding the Hankel matrix does not always yield better estimates, holds.

5 Conclusion

This paper considers an interesting question in the modelling of multivariate time series from a dual theoretical/practical perspective.

These were driven by the observation that in some papers in the context of Linear Systems, VARMA Models, etc., the dimension of certain Hankel matrices used was unnecessarily large, and the computational work can be reduced with some of the considerations proposed.

From a practical standpoint, we considered two algorithms and conducted a wide-ranging experiment using the algorithm proposed in [5] and XYW Methods in [7-11], as a result of which we can state that increasing the size of certain Hankel matrices used in some estimation methods for VARMA models does not always yield the best estimates. The accurate of the solution depends on the chosen example.

For future research it is interesting to study the influence of the noise and quantify the improvement of the estimation. Moreover, we propose conducting similar studies using any of the linear system estimation methods that depend on Hankel matrices to determine whether or not it is necessary, in practice, to consider dimensions higher than the smallest dimension that is theoretically required. Determining those specific cases where an expanded matrix could improve the estimates, and the size recommended for each particular example, will require further research.

Acknowledgments: We wish to thank Prof. J. Mari for sending us the software used in [5] and for his help with our questions on his MATLAB algorithm. The work of C. González-Concepción and C. Pestano-Gabino was supported in part by Spain's Ministerio de Economía, Industria y Competitividad MTM2015-63680-R:180.830.1602 (MINECO/FEDER) and MTM2015-71352-P, respectively.

References:

[1] H. Lütkepohl, The New Introduction to Multiple Time Series Analysis, Springer-Verlag, Berlin, 2005.

- [2] S. Nsiri and R. Roy, Identification of Refined ARMA Echelon Form Models for Multivariate Time Series, *Journal of Multivariate Analysis* 56, 207-231(1996).
- [3] R.S. Tsay, Parsimonious Parameterization of Vector Autoregressive Moving Average Models, *Journal of Business & Economic Statistics*, Vol. 7, No. 3, 327-341(1989).
- [4] G.C. Tiao and R.S. Tsay, Model Specification in Multivariate Time Series, *Journal of the Royal Statistical Society B* 51 (2): 157-213 (1989).
- [5] J. Mari, P. Stoica and T. McKelvey, Vector ARMA Estimation: A Reliable Subspace Approach, *IEEE Transactions on Signal Processing*, Vol. 48, No. 7, 2092-2104 (2000).
- [6] M. Fazel, T.K. Pong, D. Sun and P. Tseng, Hankel Matrix Rank Minimization with Applications to System Identification and Realization, *SIAM Journal Matrix Analysis and Applications*, 34(3), 946-977 (2013).
- [7] B.D. Anderson, M. Deistler, E. Felsenstein, B. Funovits, L. Koelbl and M. Zamani, Multivariate AR systems and mixed frequency data: G-identifiability and estimation, *Econometric Theory*, 32(4), 793-826 (2016).
- [8] B.D. Anderson, M. Deistler, E. Felsenstein and L. Koelbl, The structure of multivariate AR and ARMA systems: Regular and singular systems; the single and the mixed frequency case. *Journal of Econometrics*, 192(2), 366-373 (2016).
- [9] B. Chen and P.A. Zdrozny, An extended Yule-Walker method for estimating a vector autoregressive model with mixed-frequency data, *Advances in Econometrics* 13, 47-73 (1998).
- [10] P.A. Zdrozny, Extended Yule-Walker identification of VARMA models with single- or mixed-frequency data. *Journal of Econometrics*, 193(2), 438-446 (2016).
- [11] L. Koelbl, A. Braumann, E. Felsenstein and M. Deistler, Estimation of VAR Systems from Mixed-Frequency Data: The Stock and the Flow Case. In *Dynamic Factor Models*. Published online: 07 Jan 2016,43-73 (2016).
- [12] C. Pestano-Gabino, C. González-Concepción and M.C. Gil-Fariña, Sure Overall Orders to Identify Scalar Component Models, *WSEAS Transactions on Mathematics*, 2006, Vol. 5(1), 97-102.