

# A Simple Nearly Unbiased Estimator of Cross-Covariances\*

Yifan Li<sup>†</sup>

Yao Rao<sup>‡</sup>

University of Manchester

University of Liverpool

This Version: October 17, 2020

## Abstract

In this paper, we propose a simple estimator of cross-covariance matrices for a multivariate time series with an unknown mean based on a linear combination of the circular sample cross-covariance estimator. Our estimator is exactly unbiased when the data generating process follows a Vector Moving Average (VMA) model with an order less than one half of the sampling period, and is nearly unbiased if such VMA model can approximate the data generating process well. In addition, our estimator is shown to be asymptotically equivalent to the conventional sample cross-covariance estimator. Via simulation, we show that the proposed estimator can to a large extent eliminate the finite sample bias of cross-covariance estimates, while not necessarily increase the mean squared error.

*MOS Subject Classification: 62H12, 62M10.*

*Keywords: cross-covariance; bias; multivariate time series.*

---

\*We thank the editor-in-chief and two anonymous referees for their constructive and insightful comments. The usual disclaimer applies.

<sup>†</sup>Corresponding author: Alliance Manchester Business School, Booth Street W, Manchester, M15 6PB, UK. e-mail: yifan.li@manchester.ac.uk.

<sup>‡</sup>Management School, University of Liverpool, Chatham St, Liverpool, L69 7ZH, UK. email: Y.Rao@liverpool.ac.uk

# 1 Introduction

The cross-covariance is a fundamental statistic that plays a central role in multivariate analysis and has extensive applications across many disciplines ranging from engineering, physics, geostatistics, bioinformatics to finance. Among its various applications, it is of particular importance in system identification and model diagnostics. It is frequently used in the testing for the multivariate white noise property (e.g. [Hosking \(1980, 1981\)](#); [Li and McLeod \(1981\)](#); [Mahdi and Ian McLeod \(2012\)](#)), heteroscedasticity and autocorrelation robust inference of long-run covariance matrix (e.g. [Newey and West \(1987\)](#); [Andrews \(1991\)](#)), spectral density estimation ([Priestley, 1981](#)); principal component analysis and factor modelling ([Wei, 2019](#)). It is usually estimated by the sample cross-covariance estimator, which is known to be biased when the population mean is unknown. However, to the best of our knowledge, there are no papers that directly address the bias of the sample cross-covariance estimator for a vector-valued time series.

Several papers provide bias-reduction or correction methods for univariate autocovariances. In a panel data framework, [Okui \(2010, 2011\)](#) propose an estimator that is asymptotically unbiased for the within-group autocovariance. However, the number of time series also approaches infinity in their asymptotic setting, which is in stark contrast to a common time-series setting with a fixed dimension. More importantly, their inference target is still univariate in nature, as cross-covariances are not estimated. A recent paper by [Vogelsang and Yang \(2016\)](#) documents a novel approach to correct for the bias of autocovariance estimators. They show that the sample autocovariances can be written as a linear combination of population covariances through a mapping matrix  $\mathbf{A}$ . By inverting the  $h$ -th order leading principal minor of  $\mathbf{A}$ , they demonstrate that their  $\mathbf{A}$ -estimator is able to deliver exactly unbiased autocovariance estimates up to the  $h$ -th lag under the condition that the univariate time series is a moving average (MA) process of order  $h$ , and is nearly unbiased when the MA( $h$ ) structure provides a good approximation. They also show that their  $\mathbf{A}$ -estimator is asymptotically equivalent to the sample autocovariance estimator.

Inspired by the work of [Vogelsang and Yang \(2016\)](#), we propose an estimator that allows exactly unbiased estimation of cross-covariance matrices if a multivariate time series is generated by a Vector MA (VMA) model with a suitably chosen VMA order, and provides sizeable bias-reduction if the VMA structure holds approximately true. Distinct from [Vogelsang and Yang \(2016\)](#) who builds a bias-corrected estimator from conventional sample autocovariances, we start from a circular version of sample cross-covariance estimator and show that it greatly simplifies the bias-correction problem for cross-covariances. This circular design is also used in [Li \(2020\)](#) to derive a bias correction for the sample third moment estimates.

Our circular-based estimator has the following unique features: (1) In a univariate framework,

our estimator delivers comparable bias-reduction of autocovariances relative to the  $\mathbf{A}$ -estimator, but with a much simpler structure. Specifically, the  $\mathbf{A}$ -estimator requires to construct the sophisticated  $\mathbf{A}$  matrix and invert its leading principal minors, while our estimator only relies on one simple mapping matrix with no inversion involved; (2) For a suitably chosen VMA order  $h$  less than one half of the sample time span and a fixed dimension  $N$ , we can estimate bias-corrected  $(h+1)N^2$  cross-covariance parameters in one step, which is unavailable from the  $\mathbf{A}$ -estimator that only applies to a univariate setting; (3) Our estimator is asymptotically equivalent to the conventional sample cross-covariance estimators, which suggests a straightforward implementation in hypothesis tests; (4) In a univariate setting, both our circular-based estimator and the original  $\mathbf{A}$ -estimator are different special cases of a generalized  $\mathbf{A}$ -estimator, which is also unbiased when the  $\text{MA}(h)$  assumption holds.

Our simulation study shows that in a univariate setting, the three bias-corrected autocovariance estimators, namely our circular-based estimator, the original  $\mathbf{A}$ -estimator and the generalized  $\mathbf{A}$ -estimator, are exactly unbiased under the  $\text{MA}(h)$  assumption, and can to a large extent reduce the bias even when the  $\text{MA}(h)$  assumption fails. The generalized  $\mathbf{A}$ -estimator has overall the smallest MSE among the three bias-corrected estimators, while our circular-based estimator and the  $\mathbf{A}$ -estimator have very close performance. The cost to pay for the bias-reduction is a higher variance, whereas gains in MSE can be achieved when the autocovariance structure is persistent. In a multivariate setting, we confirm that our estimator is indeed effective in eliminating bias for sample cross-covariance estimates.

The rest of this paper is organized as follows: Section 2 describes the general econometric setting and defines the circular-based sample cross-covariance estimator. The multivariate cross-covariance estimator is derived in Section 3 which also includes the discussion of a generalized  $\mathbf{A}$ -estimator in the univariate case. Section 4 contains the finite sample simulation study of the proposed estimator that includes both univariate and multivariate analyses. We discuss some empirical considerations and further research topics in Section 5. Section 6 concludes. All proofs are provided in the Appendix.

## 2 Setting

On a filtered probability space in discrete time  $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P})$ , we observe a real vector-valued  $N$ -dimensional process  $Y = \{Y_t\}_{t \geq 0}$ , where  $Y_t = \{y_t^{(n)}\}_{n=1:N}'$ . Throughout this paper, we will assume that the indexing variables  $h, i, j$  and  $t$  take values in  $\mathbb{Z}$  only whenever no confusion is caused. We will also assume a fixed  $N$  throughout the paper. We impose the following assumption on  $Y$  throughout the paper:

**Assumption 1.** The time series  $Y$  is a stationary and ergodic process with finite fourth moments and satisfies the absolute summability condition  $\sum_{j=1}^{\infty} \|\Gamma_j\| < \infty$  for some matrix norm  $\|\cdot\|$ , where  $\Gamma_j = \text{E}[(Y_t - \mu)(Y_{t-j} - \mu)']$  and  $\mu = \text{E}[Y_t]$ .

The sequence of  $N$ -by- $N$  matrices  $\{\Gamma_j\}_{j \in \mathbb{Z}}$  is called the cross-covariances of  $Y$ , which is the main inference target of this paper. The above assumption is satisfied by stationary VARMA models, and is required for a meaningful estimation of the cross-covariance matrices. For a realization of  $Y$  denoted by  $\{Y_t\}_{t=1:T}$ , a frequently used estimator of  $\Gamma_j$  is the following (conventional) cross-covariance estimator:

$$\hat{\Gamma}_j = \frac{1}{T} \sum_{t=(j \vee 0)+1}^{T+(j \wedge 0)} (Y_t - \bar{Y})(Y_{t-j} - \bar{Y})', \quad |j| \in [0, T-1], \quad (2.1)$$

where  $\bar{Y}$  is the sample mean, and  $\vee$  ( $\wedge$ ) is the maximum (minimum) operator. By a standard ergodic argument, the above estimator consistently estimates  $\Gamma_j$  for any fixed  $j \in \mathbb{Z}$ . However, due to the problem of an unknown  $\mu$ , it is well-known that the above estimator is biased, and the main research interest of this paper is to provide a simple finite sample bias correction method for the estimation of  $\Gamma_j$ , which, to the best of our knowledge, is not available in the extant literature.

To this end, we firstly introduce a circular definition of the sample cross-covariances in the spirit of the circular block bootstrap of [Politis and Romano \(1992\)](#):

**Definition 1.** Given a realization of  $Y$  denoted as  $\{Y_t\}_{t=1:T}$ . For each  $|j| \in [0, T-1]$ , define the time index circularly such that  $Y_{-j} = Y_{T-j}$ . The circular sample cross-covariance is defined as:

$$\hat{\Psi}_j = \frac{1}{T} \sum_{t=1}^T (Y_t - \bar{Y})(Y_{t-j} - \bar{Y})', \quad (2.2)$$

The circular sample cross-covariance has the following interesting properties:

**Proposition 1.** Let  $\{\hat{\Psi}_j\}_{j=0:T-1}$  denote the sequence of circular sample cross-covariances constructed from  $\{Y_t\}_{t=1:T}$ . The following relations hold for  $|j| \in [1, T-1]$ :

$$\hat{\Psi}_j = \hat{\Gamma}_j + \hat{\Gamma}'_{T-j}, \quad \hat{\Psi}_j = \hat{\Psi}'_{-j} = \hat{\Psi}'_{T-j}, \quad \sum_{j=0}^{T-1} \hat{\Psi}_j = 0. \quad (2.3)$$

**Remark 1.** Note that by definition  $\hat{\Psi}_0 = \hat{\Gamma}_0$ . The difference between  $\hat{\Psi}_j$  and  $\hat{\Gamma}_j$  is subtle. The conventional estimator  $\hat{\Gamma}_j$  avoids the negative indices and uses  $T - |j|$  cross terms while  $\hat{\Psi}_j$  uses all  $T$  cross terms by cycling the observations. Asymptotically,  $\hat{\Psi}_j$  and  $\hat{\Gamma}_j$  are equivalent for any fixed  $j$  because  $\hat{\Gamma}'_{T-j}$  vanishes as  $T$  diverges under the ergodic assumption. In a finite-sample setting, however,  $\hat{\Psi}_j$  is in general inferior to  $\hat{\Gamma}_j$  for the following reasons. Firstly,  $\hat{\Psi}_j = \hat{\Gamma}_j + \hat{\Gamma}'_{T-j}$  suggests that for  $j$  large relative to  $T$ ,  $\hat{\Gamma}'_{T-j}$  can dominate  $\hat{\Gamma}_j$ , yielding an estimator of  $\hat{\Gamma}'_{T-j}$  instead of  $\hat{\Gamma}_j$ . This implies that  $\hat{\Psi}_j$  is only applicable for  $j$  small relative to  $T$  such that  $\hat{\Gamma}'_{T-j}$  is close to a zero matrix. Secondly, as the additional term  $\hat{\Gamma}'_{T-j}$  is irrelevant to the inference target ( $\Gamma_j$ ),  $\hat{\Psi}_j$  can have a larger bias and variance in comparison to  $\hat{\Gamma}_j$ . However, circular sample cross-covariance estimators are computationally more convenient by simply reordering the rows of the data matrix. More importantly, the symmetry of the circular structure greatly simplifies the computation of the finite sample bias of  $\hat{\Psi}_j$ , which is summarized in the theorem below.

**Theorem 1.** The following results hold for all  $|j| \in [1, T - 1]$ :

$$\begin{aligned} \mathbb{E}[\hat{\Psi}_0] &= \frac{T-1}{T}\Gamma_0 - \frac{1}{T}K_T, \\ \mathbb{E}[\hat{\Psi}_j] &= \frac{T-j}{T}\Gamma_j + \frac{j}{T}\Gamma'_{T-j} - \frac{1}{T}\Gamma_0 - \frac{1}{T}K_T, \end{aligned} \quad (2.4)$$

where  $K_T = \sum_{j=1}^{T-1} \frac{T-j}{T}(\Gamma_j + \Gamma'_j)$  is an  $N$ -by- $N$  symmetric matrix.

**Remark 2.** We see that the bias of  $\hat{\Psi}_j$  depends on  $K_T/T$  which collects all the cross-covariances to the end of the dataset. Therefore, the bias can be sizeable when  $T$  is small, but converges to zero as  $T \rightarrow \infty$  under the absolute summability condition. Also, for  $j$  large relative to  $T$ , the bias of  $\hat{\Psi}_j$  can be substantial due to the term  $\frac{j}{T}\Gamma'_{T-j}$ , which corroborates our discussions in Remark 1.

Note that the above simple result is fully attributed to the circular structure, as an analogous result for conventional sample cross-covariances is much more difficult to derive. For example, in the supporting document of [Vogelsang and Yang \(2016\)](#), they provide a lengthy derivation of the bias for the conventional sample autocovariances, which involves a sophisticated  $\mathbf{A}$  matrix. It is also not clear whether such a result can be directly generalized to our  $N$ -dimensional setting.

To illustrate further on the estimation of  $\Gamma_j$ , we introduce some notations. Let us denote the  $j$ -th order circular sample cross-covariance between the  $m$ -th and the  $n$ -th time series as  $\hat{\Psi}_j(m, n)$ , which is simply the element on the intersection of the  $m$ -th row and  $n$ -th column of  $\hat{\Psi}_j$ :

$$\hat{\Psi}_j(m, n) = \frac{1}{T} \sum_{t=1}^T (y_t^{(m)} - \bar{y}^{(m)})(y_{t-j}^{(n)} - \bar{y}^{(n)}), \quad (2.5)$$

where  $\bar{y}^{(n)}$  denotes the sample mean of  $y_t^{(n)}$ . The terms  $K_T(m, n)$  and  $\Gamma_j(m, n)$  follow the same logic, and it is clear that we have  $\hat{\Psi}_j(m, n) = \hat{\Psi}_{T-j}(n, m)$ ,  $\hat{\Psi}_0(m, n) = \hat{\Psi}_0(n, m)$ , and  $K_T(m, n) = K_T(n, m)$  for any  $m, n \in [1, N]$ . Now consider the linear system of equations for all  $j \in [1, T - 1]$ :

$$\mathbb{E}[\hat{\Psi}_j(m, n)] = \frac{T-j}{T}\Gamma_j(m, n) + \frac{j}{T}\Gamma_{T-j}(n, m) - \frac{1}{T}\Gamma_0(m, n) - \frac{1}{T}K_T(m, n), \quad (2.6)$$

We can write the above system of equations in a compact matrix format:

$$\mathbb{E}[\underbrace{\hat{\Psi}(m, n)}_{(T-1) \times 1}] = \underbrace{\mathbf{G}}_{(T-1) \times (2T-1)} \underbrace{\mathbf{\Gamma}(m, n)}_{(2T-1) \times 1}, \quad (2.7)$$

where:

$$\hat{\Psi}(m, n) = \begin{bmatrix} \hat{\Psi}_1(m, n) \\ \vdots \\ \hat{\Psi}_{T-1}(m, n) \end{bmatrix}, \quad \mathbf{\Gamma}(m, n) = \begin{bmatrix} \Gamma_0(m, n) \\ \Gamma_1(m, n) \\ \vdots \\ \Gamma_{T-1}(m, n) \\ \Gamma_1(n, m) \\ \vdots \\ \Gamma_{T-1}(n, m) \end{bmatrix}, \quad (2.8)$$

and the matrix  $\mathbf{G} = \{G_{i,j}\}_{i=1:T-1, j=1:2T-1}$  is defined as follows:

$$G_{i,j} = \begin{cases} -\frac{1}{T}, & i \in [1, T-1], j = 1 \\ \frac{T-j+1}{T} \left( \mathbb{1}_{\{i+1=j\}} - \frac{1}{T} \right), & i \in [1, T-1], j \in [2, T] \\ G_{T-i, j-T+1}, & i \in [1, T-1], j \in [T+1, 2T-1] \end{cases}, \quad (2.9)$$

where  $\mathbb{1}_A$  is the indicator function which equals 1 if the set  $A$  is non-empty, and zero otherwise. Note that we deliberately exclude  $\hat{\Psi}_0(m, n)$  in the construction of  $\hat{\Psi}(m, n)$  due to  $\sum_{j=0}^{T-1} \hat{\Psi}_j(m, n) = 0$  from Proposition 1, so it does not add any information to the system of equations. It is also clear that since  $\hat{\Psi}_{T-j}(n, m) = \hat{\Psi}_j(m, n)$  from Proposition 1, we only need to include either  $\hat{\Psi}(m, n)$  or  $\hat{\Psi}(n, m)$ . We are now in a classic linear Generalized Method of Moments (GMM) setting, where an unbiased estimator of  $\mathbf{\Gamma}(m, n)$  can be formed by finding a  $(2T-1)$ -by- $(T-1)$  matrix  $\mathbf{R}$  such that:

$$\mathbf{R}\mathbf{E}[\hat{\Psi}(m, n)] = \mathbf{\Gamma}(m, n), \quad (2.10)$$

or equivalently  $\mathbf{R}\mathbf{G} = \mathbf{I}_{2T-1}$ , where  $\mathbf{I}_n$  is the  $n$ -by- $n$  identity matrix. However, in our setting, such an  $\mathbf{R}$  does not exist because the  $\mathbf{G}$  matrix does not have full column rank, thus the system is under-identified. To illuminate this problem, let us consider two different scenarios.

**Case 1:**  $m = n$ . In this case, we have  $\hat{\Psi}_j(m, m) = \hat{\Psi}_{T-j}(m, m)$  for all  $j \in [1, T-1]$ . Therefore, there are  $\lfloor \frac{T}{2} \rfloor$  unique sample moments which means that only  $\lfloor \frac{T}{2} \rfloor$  moments from  $\mathbf{\Gamma}(m, m)$  can be identified. Suppose that  $\Gamma_0(m, m)$  is always estimated, then we can at most identify  $\lfloor \frac{T}{2} \rfloor - 1$  autocovariance terms from  $\{\Gamma_j(m, m)\}_{j=1:T-1}$ . Assuming that the autocovariance structure of a time series decays to zero as  $j$  expands, a reasonable choice in practice is to identify  $\Gamma_j(m, m)$  for  $j \in [1, \lfloor \frac{T}{2} \rfloor - 1]$  and assume that  $\Gamma_j(m, m) = 0$  for  $j \geq \lfloor \frac{T}{2} \rfloor$ .

**Case 2:**  $m \neq n$ . In this case, all sample moments in  $\hat{\Psi}_j(m, n)$  are unique, thus at most  $T-1$  moments can be identified from  $\mathbf{\Gamma}(m, n)$ . Suppose that  $\Gamma_0(m, n) = \Gamma_0(n, m)$  is always estimated and we would like to estimate an equal amount of lags from  $\{\Gamma_j(m, n)\}_{j=1:T-1}$  and  $\{\Gamma_j(n, m)\}_{j=1:T-1}$ , then it is also clear that at most  $\lfloor \frac{T}{2} \rfloor - 1$  lags of cross-covariances can be identified in this case. From a similar argument as in Case 1, we choose to identify  $\Gamma_j(n, m)$  and  $\Gamma_j(m, n)$  for  $j \in [1, \lfloor \frac{T}{2} \rfloor - 1]$  and assume the remainder terms to be zero.

Based on the above discussion, it is clear that identification restrictions are required for such an  $\mathbf{R}$  to exist. Following the idea in [Vogelsang and Yang \(2016\)](#), we assume that  $Y$  follows a VMA( $h$ ) model, formally defined as below:

**Definition 2.** The process  $Y$  is generated by a VMA( $h$ ) model with a fixed  $h < \infty$  if  $Y$  has the following representation:

$$Y_t = \mu + \sum_{i=0}^h \Theta_i \mathbf{u}_{t-i}, \quad (2.11)$$

where  $\Theta_i$  are  $N$ -by- $N$  real coefficient matrices and  $\mathbf{u}_t$  is a vector-valued weak white noise process.

Under the VMA( $h$ ) assumption, it is clear that  $\Gamma_j = \mathbf{0}_{N \times N}$  for all  $j > h$ , which provides identification restrictions for our estimation problem stated earlier. We are now ready to state the main results in our paper.

### 3 Main Results

Our main result is the following theorem:

**Theorem 2.** Assume that  $Y$  follows a VMA( $h$ ) model with a fixed  $h < \lfloor \frac{T}{2} \rfloor$  as in Definition 2. Define the  $(T-1)$ -by- $(2h+1)$  matrix  $\mathbf{H}_h = \{H_{i,j}\}_{i=1:T-1, j=1:2h+1}$  as:

$$H_{i,j} = \begin{cases} -\frac{1}{T}, & i \in [1, T-1], j = 1 \\ \frac{T-j+1}{T} \left( \mathbb{1}_{\{i+1=j\}} - \frac{1}{T} \right), & i \in [1, T-1], j \in [2, h+1], \\ H_{T-i, j-h}, & i \in [1, T-1], j \in [h+2, 2h+1]. \end{cases} \quad (3.1)$$

Let  $\mathbf{W}$  denote a  $(T-1)$ -by- $(T-1)$  non-random positive definite matrix, then for all  $m, n \in [1, N]$ , we can construct the following estimator:

$$\underbrace{\widehat{\Psi}_h^*(m, n)}_{(2h+1) \times 1} \equiv \underbrace{(\mathbf{H}'_h \mathbf{W} \mathbf{H}_h)^{-1} \mathbf{H}'_h \mathbf{W}}_{(2h+1) \times (T-1)} \underbrace{\widehat{\Psi}(m, n)}_{(T-1) \times 1}. \quad (3.2)$$

For all  $m, n \in [1, N]$  and valid choices of  $\mathbf{W}$ , it holds that:

$$\mathbb{E}[\widehat{\Psi}_h^*(m, n)] = \mathbf{\Gamma}_h, \quad \widehat{\Psi}_h^*(m, n) \xrightarrow{P} \mathbf{\Gamma}_h(m, n), \quad (3.3)$$

where  $\mathbf{\Gamma}_h(m, n) = \{\Gamma_0(m, n), \Gamma_1(m, n), \dots, \Gamma_h(m, n), \Gamma_1(n, m), \dots, \Gamma_h(n, m)\}'$  is the  $(2h+1)$ -by-1 vector collecting all the cross-covariances terms between the  $m$ -th and the  $n$ -th series up to lag  $h$ .

**Remark 3.** By construction,  $\mathbf{H}_h$  is a sub-matrix of  $\mathbf{G}$  that removes the columns representing  $\Gamma_j(m, n)$  and  $\Gamma_j(n, m)$  for  $j > h$ . In fact we have  $\mathbf{H}_{T-1} = \mathbf{G}$ . The upper bound of  $h$  is a result from our identification restriction as discussed in the previous section. In detail,  $\mathbf{H}_h$  has full column rank for all  $h < \lfloor \frac{T}{2} \rfloor$ , so that the linear GMM estimator is at least just-identified. As the above result holds for any  $m$  and  $n$ , one can construct bias-corrected estimates for all elements in  $\{\Gamma_j\}_{j=0:h}$  by computing  $\widehat{\Psi}_h^*(m, n)$  for all  $m$  and  $n$ . Throughout this paper, the superscript  $*$  denotes a bias-corrected estimator.

**Remark 4.** The truncation parameter  $h$  plays a similar role as the parameter  $m$  in the  $\mathbf{A}$ -estimator (a definition is given in Section 3.1), which controls for the number of cross-covariance terms to be estimated. In a finite sample,  $h$  can take any value below  $\lfloor \frac{T}{2} \rfloor$ , while Vogelsang and Yang (2016) claim that  $m$  can be as large as  $T-2$ . Therefore for autocovariance estimation, the  $\mathbf{A}$ -estimator can potentially correct for the bias caused by a more persistent dynamic structure than our approach, and

can be more efficient as  $\hat{\Gamma}_j$  is superior to  $\hat{\Psi}_j$ . However, as will be clear from the results below, our estimator is much simpler to construct and is directly applicable to the multivariate case. Moreover, in our simulation section, we demonstrate that the performance of our estimator is almost as precise as the  $\mathbf{A}$ -estimator in a univariate setting.

We also have the following result which reconciles our estimator with the bias-corrected sample variance-covariance estimator under the i.i.d. assumption:

**Corollary 1.** Suppose  $h = 0$  and  $\mathbf{W} = \mathbf{I}_{T-1}$ , then it holds that:

$$\hat{\Psi}_0^*(m, n) = \frac{T}{T-1} \hat{\Psi}_0(m, n). \quad (3.4)$$

Theorem 2 guarantees that  $\hat{\Psi}_h^*(m, n)$  will be nearly unbiased when  $\text{VMA}(h)$  is a good approximation to the covariance structure of  $y$ . In fact, it is easy to derive the exact bias of  $\hat{\Psi}_h^*(m, n)$  when the  $\text{VMA}(h)$  assumption does not hold.

**Proposition 2.** Suppose  $Y$  follows a  $\text{VMA}(\infty)$  process, then it holds that:

$$\mathbb{E}[\hat{\Psi}_h^*(m, n)] = \Gamma_h(m, n) + (\mathbf{H}'_h \mathbf{W} \mathbf{H}_h)^{-1} \mathbf{H}'_h \mathbf{W} \boldsymbol{\epsilon}(m, n), \quad (3.5)$$

where  $\boldsymbol{\epsilon}(m, n) = \{\epsilon_j(m, n)\}_{j=1:T-1}$  and:

$$\begin{aligned} \epsilon_j(m, n) = & -\frac{1}{T} \sum_{k=h+1}^{T-1} \frac{T-k}{T} (\Gamma_k(m, n) + \Gamma_k(n, m)) \\ & + \mathbb{1}_{\{j < T-h\}} \frac{j}{T} \Gamma_{T-j}(n, m) + \mathbb{1}_{\{j > h\}} \frac{T-j}{T} \Gamma_j(m, n). \end{aligned} \quad (3.6)$$

**Remark 5.** Intuitively,  $\boldsymbol{\epsilon}(m, n)$  contains all of the cross-covariance terms which are assumed to be zero under the  $\text{VMA}(h)$  assumption. Note that the first term in  $\boldsymbol{\epsilon}(m, n)$  is only of the order  $\mathcal{O}(T)$  which is negligible compared to the other two terms. Suppose  $\Gamma_j(m, n)$  decays exponentially, then as long as  $h$  is chosen such that  $\Gamma_j(m, n) \approx \Gamma_j(n, m) = 0$  for  $j > h$ ,  $\hat{\Psi}_h^*(m, n)$  will be nearly unbiased even when the  $\text{VMA}(h)$  assumption fails.

To construct  $\hat{\Psi}_h^*(m, n)$  in practice, we still have to choose  $\mathbf{W}$ , which can be different for each  $m$  and  $n$ . In fact, one can even choose an adaptive  $\mathbf{W}$  for every  $m$  and  $n$ , which minimizes the variance of the estimator. However, an adaptive  $\mathbf{W}$  removes the unbiasedness property of  $\hat{\Psi}_h^*(m, n)$ , and is thus not considered in this paper. Even if we keep  $\mathbf{W}$  fixed, we still need to compute  $(\mathbf{H}'_h \mathbf{W} \mathbf{H}_h)^{-1} \mathbf{H}'_h \mathbf{W}$ , which involves a matrix inversion that can be undesirable in the case of large  $N$ .

To solve the aforementioned implementation issues, we derive an equivalent representation of  $\hat{\Psi}_h^*(m, n)$  based on the identity weighting matrix  $\mathbf{W} = \mathbf{I}_{T-1}$ , which can be computed explicitly without any matrix inversion. As will be shown later, this choice actually implies an asymptotic equivalence between  $\hat{\Psi}_h^*(m, n)$  and  $\hat{\Gamma}_j(m, n)$ .



**Proposition 3.** Set  $\mathbf{W} = \mathbf{I}_{T-1}$ , then for every  $m$  and  $n$ , it holds that  $\widehat{\Psi}_h^*(m, n) = \mathbf{R}_h \widehat{\Psi}(m, n)$ , where  $\mathbf{R}_h$  for  $0 < h < \lfloor \frac{T}{2} \rfloor$  is defined as:

$$\underbrace{\mathbf{R}_h}_{(2h+1) \times (T-1)} = \begin{bmatrix} -1, & -1, & \dots, & -1, & -(T-2h)\mathbf{v}'_h, & -1, & \dots, & -1, & -1 \\ \frac{T}{T-1}, & 0, & \dots, & 0, & -\frac{T}{T-1}\mathbf{v}'_h, & 0, & 0, & 0, & 0 \\ 0, & \frac{T}{T-2}, & \dots, & 0, & -\frac{T}{T-2}\mathbf{v}'_h, & 0, & 0, & 0, & 0 \\ \vdots, & \vdots, & \ddots, & \vdots, & \vdots, & \vdots, & \vdots, & \vdots, & \vdots \\ 0, & 0, & \dots, & \frac{T}{T-h}, & -\frac{T}{T-h}\mathbf{v}'_h, & 0, & 0, & 0, & 0 \\ 0, & 0, & 0, & 0, & -\frac{T}{T-1}\mathbf{v}'_h, & 0, & 0, & \dots, & \frac{T}{T-1} \\ 0, & 0, & 0, & 0, & -\frac{T}{T-2}\mathbf{v}'_h, & 0, & \dots, & \frac{T}{T-2}, & 0 \\ \vdots, & \vdots, & \vdots, & \vdots, & \vdots, & \vdots, & \ddots, & \vdots, & \vdots \\ 0, & 0, & 0, & 0, & -\frac{T}{T-h}\mathbf{v}'_h, & \frac{T}{T-h}, & 0, & \dots, & 0 \end{bmatrix}, \quad (3.7)$$

and  $\mathbf{v}_h$  is the  $(T-2h-1)$ -by-1 equal weight vector (with all elements equal and summing up to 1). Specially,  $\mathbf{R}_0 = -T\mathbf{v}'_0$ .

Proposition 3 allows us to estimate any selected elements of  $\mathbf{\Gamma}_h(m, n)$  by taking the corresponding rows from  $\mathbf{R}_h$  and multiplying it with  $\widehat{\Psi}(m, n)$ . In fact, we can even estimate all  $(h+1)N^2$  parameters from  $\{\Gamma_j\}_{j=0:h}$  in one step using the following result:

**Corollary 2.** Let  $\text{vec}(\cdot)$  denote the vectorization operator which stacks all elements of a matrix into a column vector (see e.g. [Henderson and Searle \(1979\)](#) for a detailed definition). Construct the matrices  $\widehat{\Psi}^\circ$  and  $\mathbf{\Gamma}_h^\circ$  as follows:

$$\underbrace{\widehat{\Psi}^\circ}_{(T-1) \times N^2} = \begin{bmatrix} \text{vec}(\widehat{\Psi}_1)' \\ \vdots \\ \text{vec}(\widehat{\Psi}_{T-1})' \end{bmatrix}, \quad \underbrace{\mathbf{\Gamma}_h^\circ}_{(2h+1) \times N^2} = \begin{bmatrix} \text{vec}(\Gamma_0)' \\ \text{vec}(\Gamma_1)' \\ \vdots \\ \text{vec}(\Gamma_h)' \\ \text{vec}(\Gamma'_1)' \\ \vdots \\ \text{vec}(\Gamma'_h)' \end{bmatrix} \quad (3.8)$$

Let us denote  $\widehat{\Psi}_h^* = \mathbf{R}_h \widehat{\Psi}^\circ$ , then under the VMA( $h$ ) assumption of  $Y$  with a fixed  $h < \lfloor \frac{T}{2} \rfloor$ . It holds that:

$$\mathbb{E}[\widehat{\Psi}_h^*] = \mathbf{\Gamma}_h^\circ, \quad \widehat{\Psi}_h^* \xrightarrow{P} \mathbf{\Gamma}_h^\circ. \quad (3.9)$$

The above results demonstrate the simplicity and computational advantage of our estimator, as we can construct a bias-corrected estimator for any sub-matrices of  $\mathbf{\Gamma}_h^\circ$  just by selecting an appropriate range of  $\mathbf{R}_h$  and  $\widehat{\Psi}^\circ$ , both of which can be constructed explicitly. We also have the following alternative representation for estimating a specific  $\Gamma_j$  as a matrix:

**Corollary 3.** Define the estimator  $\hat{\Psi}_{j,h}^*$  as follows:

$$\hat{\Psi}_{j,h}^* = \frac{T}{T-j} \left( \hat{\Psi}_j + \frac{1}{T-2h-1} \sum_{i=-h}^h \hat{\Psi}_i \right), \quad 0 \leq |j| \leq h. \quad (3.10)$$

Under the VMA( $h$ ) assumption of  $Y$  with a fixed  $h < \lfloor \frac{T}{2} \rfloor$ , we have for all  $j \leq h$ :

$$\mathbb{E}[\hat{\Psi}_{j,h}^*] = \Gamma_j, \quad \hat{\Psi}_{j,h}^* \xrightarrow{P} \Gamma_j \quad (3.11)$$

**Remark 6.** In view of Proposition 2, the bias of  $\hat{\Psi}_{j,h}^*$  can be derived explicitly using the  $\mathbf{R}_h$  matrix after some simple calculation:

$$\mathbb{E}[\hat{\Psi}_{j,h}^*] - \Gamma_j = \frac{j}{T-j} \Gamma_{T-j} - \frac{T}{(T-2h-1)(T-j)} \sum_{k=h+1}^{T-h-1} \frac{T-k}{T} (\Gamma_k + \Gamma'_k), \quad j \in [0, h]. \quad (3.12)$$

Therefore, this bias is in general very small if  $\Gamma_j$  is close to be a zero matrix for all  $j > h$ , and it is indeed of order  $\mathcal{O}(T^{-1})$  for any fixed  $h$ , which follows from the consistency of  $\hat{\Psi}_{j,h}^*$ . Comparing with the expression in Eq. (2.6), we should expect the bias of  $\hat{\Psi}_{j,h}^*$  to be much smaller than that of  $\hat{\Psi}_j$  when the low order cross-covariances dominates the high order ones, because the bias of  $\hat{\Psi}_{j,h}^*$  only consists of  $\Gamma_j$  for  $j > h$ . In practice, one may plug  $\hat{\Gamma}_j$  into the above formula to approximate the bias of  $\hat{\Psi}_{j,h}^*$ , however it should be noted that such approximation is in itself biased, which does not necessarily reflect the actual bias of the estimator.

An important property of a variance-covariance estimator is that it should be positive definite. We discuss the positive definiteness of  $\hat{\Psi}_{0,h}^*$  in the result below:

**Proposition 4.** Based on a sample size of  $T$  and some  $h < \lfloor \frac{T}{2} \rfloor$ ,  $\hat{\Psi}_{0,h}^*$  is almost surely positive definite if it satisfies the following criterion:

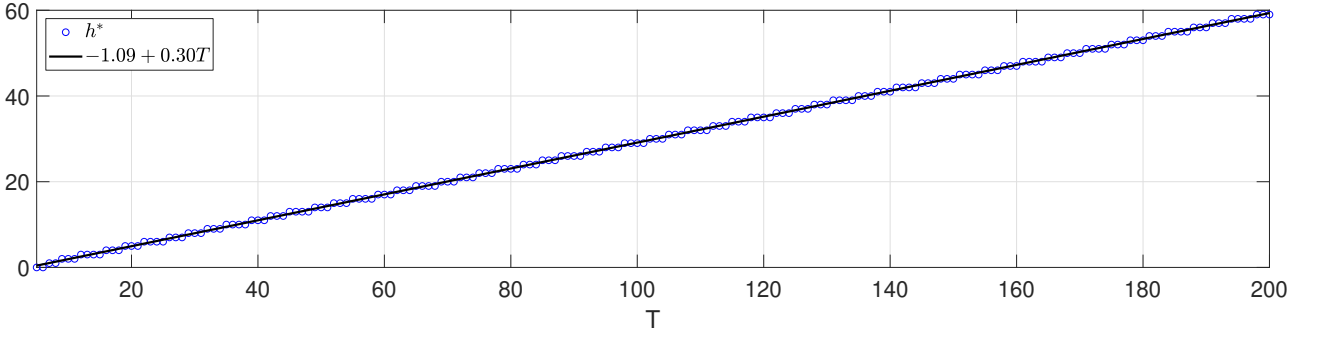
$$\inf_{x \in [-\pi, \pi]} 2\pi(D_{T-1}(x) - D_{T-h-1}(x) + D_h(x)) > 1 + 2h - T, \quad (3.13)$$

where  $D_h(x) = \frac{\sin((2h+1)x/2)}{2\pi \sin(x/2)}$  is the  $h$ -th order Dirichlet kernel.

**Remark 7.** The criterion in Eq. (3.13) can be easily checked for any  $T$  and  $h$  numerically. Using the crude bound  $|D_h(x)| \leq \frac{2h+1}{2\pi}$ , it is easy to see that  $\hat{\Psi}_{0,h}^*$  must be almost surely positive definite for all  $h \leq \frac{T-1}{6}$ . We are unable to sharpen this bound analytically. Instead, for each  $T$ , we calculate the largest  $h^*$  such that Eq. (3.13) is satisfied for all  $h \leq h^*$ , which is presented in Figure 1. The figure shows that  $h^*$  can be very well approximated by  $-1.09 + 0.30T$ , which provides a sharp upper bound for the choice of  $h$  in practice that guarantees a positive definite  $\hat{\Psi}_{0,h}^*$ . This, inevitably, further restricts the range of  $h$ , which is a problem shared by the  $\mathbf{A}$ -estimator if the strict positivity of variance estimates is pursued.

Finally, we show that our cross-covariance estimator is asymptotically equivalent to the conventional sample cross-covariance estimator:

Figure 1: Numerically computed  $h^*$  for  $T \in [5, 200]$



Note: For each  $T \in [5, 200]$ , we compute numerically the largest  $h^*$  such that Eq. (3.13) is satisfied for all  $h \leq h^*$ . The black line represents the linear least square fit to the  $h^*$ s.

**Theorem 3.** Assume that  $Y$  is a  $VMA(\infty)$  process satisfying the conditions in Assumption 1. For fixed integers  $N$  and  $h$ , it holds that:  $\sqrt{T}(\widehat{\Psi}_h^* - \widehat{\Gamma}_h^\circ) = o_p(1)$ , where  $\widehat{\Gamma}_h^\circ$  is the conventional sample cross-covariance estimator of  $\Gamma_h^\circ$ .

**Remark 8.** Note that the above result implies that  $\widehat{\Psi}_h^*(m, m)$  is asymptotically equivalent to the  $\mathbf{A}$ -estimator of Vogelsang and Yang (2016), since the  $\mathbf{A}$ -estimator is also asymptotically equivalent to the sample autocovariance estimator. Therefore, one can use our estimator in tests by simply replacing the conventional sample cross-covariance estimators, and it is likely to improve the finite sample performance of the original test, which is supported by the findings in Vogelsang and Yang (2016); Yang and Vogelsang (2018).

### 3.1 Connection to the $\mathbf{A}$ -Estimator in the Univariate Case

As both our estimator and the  $\mathbf{A}$ -estimator of Vogelsang and Yang (2016) are unbiased in a univariate setting, it is important to understand the relationship between these two estimators. To this end, we firstly state the definition of the  $\mathbf{A}$ -estimator. Let us denote  $\gamma = \{\gamma_j\}'_{j=0:T-1}$  as the  $T$ -by-1 vector of true autocovariances for a univariate time series  $y$ , and denote  $\widehat{\gamma} = \{\widehat{\gamma}_j\}'_{j=0:T-1}$  denote the  $T$ -by-1 vector of conventional sample variance and autocovariances. For some  $h \leq T - 2$ , the  $\mathbf{A}$ -estimator  $\widetilde{\gamma}_h^*$  is defined as:

$$\underbrace{\widetilde{\gamma}_h^*}_{(h+1) \times 1} = \underbrace{\mathbf{A}_h^{-1}}_{(h+1) \times (h+1)} \underbrace{\widehat{\gamma}_h}_{(h+1) \times 1}, \quad (3.14)$$

where  $\widehat{\gamma}_h = \{\widehat{\gamma}_j\}'_{j=0:h}$ ,  $\mathbf{A}_h$  is the leading principle minor of order  $h + 1$  of the  $T$ -by- $T$  matrix  $\mathbf{A}$  which satisfies the property  $E[\widehat{\gamma}] = \mathbf{A}\gamma$ . Its expression can be found in Section 3 of Vogelsang and Yang (2016). It then follows that  $E[\widetilde{\gamma}_h^*] = \gamma_h$  with  $\gamma_h = \{\gamma_j\}'_{j=0:h}$  if  $y$  follows an  $MA(h)$  model, and  $\widetilde{\gamma}_h^*$  is asymptotically equivalent to  $\widehat{\gamma}_h$ . Note that here  $h$  is understood as the parameter  $m$  in Vogelsang and Yang's (2016) original notation.

For some  $h < \lfloor \frac{T}{2} \rfloor$ , our univariate bias-corrected autocovariance estimator takes the following form due to Proposition 3:

$$\underbrace{\widehat{\boldsymbol{\psi}}_h^*}_{(h+1) \times 1} = \underbrace{\widetilde{\mathbf{R}}_h}_{(h+1) \times (T-1)} \underbrace{\widehat{\boldsymbol{\psi}}}_{(T-1) \times 1}, \quad (3.15)$$

where  $\widehat{\boldsymbol{\psi}} = \{\widehat{\psi}_j\}'_{j=1:T-1}$  is the  $(T-1)$ -by-1 vector of circular sample autocovariances, and  $\widetilde{\mathbf{R}}_h$  is defined as the first  $h+1$  rows of  $\mathbf{R}_h$  in Eq. (3.7). Intuitively the last  $h$  rows of  $\mathbf{R}_h$  are not needed due to the fact that  $\gamma_j = \gamma_{-j}$  by definition. To compare our estimator with  $\widetilde{\boldsymbol{\gamma}}_h^*$ , we present the following alternative representation of  $\widehat{\boldsymbol{\psi}}_h^*$  in terms of conventional sample autocovariances:

**Corollary 4.** The following result holds:

$$\underbrace{\widehat{\boldsymbol{\psi}}_h^*}_{(h+1) \times 1} = \underbrace{\mathbf{M}_h}_{(h+1) \times (2h+1)} \underbrace{\widehat{\boldsymbol{\gamma}}_J}_{(2h+1) \times 1}, \quad (3.16)$$

where  $\widehat{\boldsymbol{\gamma}}_J = \{\widehat{\gamma}_j\}_{j \in J}$ ,  $J = \{0, 1, \dots, h, T-h, \dots, T-1\}$  and  $\mathbf{M}_h = \{M_{i,j}\}_{i=1:h+1, j=1:2h+1}$  has the following form:

$$M_{i,j} = \begin{cases} \frac{T}{T-i+1} (\mathbb{1}_{\{i=1\}} + \frac{1}{T-2h-1}), & i \in [1, h+1], j = 1 \\ \frac{T}{T-i+1} (\mathbb{1}_{\{i=j\} \cup \{i=T-j+1\}} + \frac{2}{T-2h-1}), & i \in [1, h+1], j \in [2, 2h+1] \end{cases} \quad (3.17)$$

Specially,  $\widetilde{\mathbf{M}}_0 = \frac{T}{T-1}$ .

**Remark 9.** The proposition reveals that  $\widehat{\boldsymbol{\psi}}_h^*$  is in general very different from  $\widetilde{\boldsymbol{\gamma}}_h^*$ . Specifically,  $\widehat{\boldsymbol{\psi}}_h^*$  uses  $2h+1$  conventional sample moments  $\{\widehat{\gamma}_j\}_{j \in J}$  to identify  $h+1$  population moments, while  $\widetilde{\boldsymbol{\gamma}}_h^*$  only uses  $h+1$  conventional sample moments  $\{\widehat{\gamma}_j\}_{j=0:h}$  to identify  $h+1$  population moments. Therefore,  $\widehat{\boldsymbol{\psi}}_h^*$  corresponds to a particular solution of an over-identified system, whereas  $\widetilde{\boldsymbol{\gamma}}_h^*$  is the unique solution of a just-identified system. As a result,  $\widehat{\boldsymbol{\psi}}_h^*$  and  $\widetilde{\boldsymbol{\gamma}}_h^*$  are never identical unless both estimators are based on the same sets of sample moments with just-identified population moments, which only holds when  $h = 0$  with the index set  $j \in J = \{0\}$ .

In the spirit of Theorem 2, we can generalize the  $\mathbf{A}$ -estimator to construct an unbiased autocovariance estimator using more than  $h$  conventional sample autocovariances, which nests both  $\widetilde{\boldsymbol{\gamma}}_h^*$  and  $\widehat{\boldsymbol{\psi}}_h^*$ . The generalized  $\mathbf{A}$ -estimator is defined as follows:

**Proposition 5.** For any fixed  $h \leq T-2$ , choose an integer  $q \in [h+1, T]$ , and let  $J_q \subseteq \{0, \dots, T-1\}$  denote an increasing sequence of indices with  $q$  terms. Denote the vector  $\widehat{\boldsymbol{\gamma}}_{J_q} = \{\widehat{\gamma}_j\}'_{j \in J_q}$  as the  $q$ -by-1 vector of conventional autocovariances, and define the  $q$ -by- $(h+1)$  matrix  $\mathbf{A}_{J_q}$  as the submatrix of  $\mathbf{A}$  with row index given by  $J_q$  and the last  $T-h-1$  columns removed. Also, let  $\mathbf{W}$  denote an arbitrary  $q$ -by- $q$  non-random positive definite matrix. Assume that  $\mathbf{A}_{J_q}$  has full column rank and that  $y$  is an MA( $h$ ) process, the generalized  $\mathbf{A}$ -estimator defined as:

$$\underbrace{\widehat{\boldsymbol{\gamma}}_{h, J_q}^*}_{(h+1) \times 1} \equiv \underbrace{(\mathbf{A}'_{J_q} \mathbf{W} \mathbf{A}_{J_q})^{-1} \mathbf{A}'_{J_q} \mathbf{W}}_{(h+1) \times q} \underbrace{\widehat{\boldsymbol{\gamma}}_{J_q}}_{q \times 1}, \quad (3.18)$$

is an unbiased and consistent estimator of  $\gamma_h$  in the sense that:

$$\mathbb{E}[\widehat{\gamma}_{h,J_q}^*] = \gamma_h, \quad \widehat{\gamma}_{h,J_q}^* \xrightarrow{P} \gamma_h. \quad (3.19)$$

**Remark 10.** It turns out that  $\widetilde{\gamma}_h^*$  is just a special case of  $\widehat{\gamma}_{h,J_q}^*$  by setting  $q = h + 1$  and  $J_q = [0, h]$  with arbitrary  $\mathbf{W}$ . In view of Corollary 4,  $\widehat{\psi}_h^*$  is also a special case of  $\widehat{\gamma}_{h,J_q}^*$  with  $h < \lfloor \frac{T}{2} \rfloor$ ,  $q = 2h + 1$ ,  $J_q = J$  and any positive definite  $\mathbf{W}$  satisfying the matrix equation:

$$\mathbf{A}'_{J_q} \mathbf{W} (\mathbf{I}_q - \mathbf{A}'_{J_q} \mathbf{M}_h) = \mathbf{0}_{(h+1) \times q}. \quad (3.20)$$

This further illustrate our previous point that  $\widehat{\psi}_h^*$  and  $\widetilde{\gamma}_h^*$  are indeed different special cases of the generalized  $\mathbf{A}$ -estimator, and the use of circular sample autocovariances in  $\widehat{\psi}_h^*$  is equivalent to a specific constraint placed on  $\mathbf{W}$ . Due to the complicated structure of  $\mathbf{A}$ , we are unable to provide an explicit form of  $\mathbf{W}$  that satisfies the above equation. However, solving for such  $\mathbf{W}$  is empirically unimportant as  $\widehat{\psi}_h^*$  can always be constructed explicitly based on Eq. (3.15) or Eq. (3.16).

**Remark 11.** Note that the matrix  $\mathbf{A}_{J_q}$  is of full column rank for any  $h \leq T - 2$  and  $q \in [h + 1, T]$  whenever  $\text{rank}(\mathbf{A}) = T - 1$ . Although Vogelsang and Yang (2016) did not prove this result analytically, their numerical experiment shows that it holds for a large range of  $T \geq 2$ . To provide more insights into the rank of  $\mathbf{A}$ , we conjecture that the reduced row echelon form (rref) of  $\mathbf{A}$  has the following simple form:

$$\text{rref}(\mathbf{A}) = \begin{bmatrix} \mathbf{I}_{T-1}, & -\boldsymbol{\nu}_{(T-1) \times 1} \\ \mathbf{0}'_{(T-1) \times 1}, & 0 \end{bmatrix}_{T \times T}, \quad (3.21)$$

where  $\boldsymbol{\nu}_{m \times n}$  is an  $m$ -by- $n$  matrix of ones. The above result can be easily verified by standard statistical software which directly implies  $\text{rank}(\mathbf{A}) = T - 1$  for arbitrary  $T \geq 2$ . However, we are unable to derive an analytical proof of this statement, and is left for future research.

We point out that for the generalized  $\mathbf{A}$ -estimator,  $q$  can take any value from  $[h + 1, T]$ . Therefore, one might expect that choosing  $q = T$  can further improve the performance of the original  $\mathbf{A}$ -estimator by involving all available sample moments. In our simulation we verify this claim by examining the finite sample performance of the generalized  $\mathbf{A}$ -estimator using all moment conditions with  $\mathbf{W} = \mathbf{I}_T$ , which has the following form:

$$\widehat{\gamma}_{h,J_T}^* = (\mathbf{A}'_{J_T} \mathbf{A}_{J_T})^{-1} \mathbf{A}'_{J_T} \widehat{\gamma}. \quad (3.22)$$

Finite sample bias and variance of  $\widehat{\gamma}_{h,J_T}^*$  can be derive in a similar fashion based on the results of  $\widehat{\psi}_h^*$  and the discussion in Vogelsang and Yang (2016), and is omitted for brevity. We conclude this section by establishing the asymptotic equivalence between  $\widehat{\gamma}_{h,J_T}^*$  and  $\widehat{\gamma}_h$ :

**Theorem 4.** Suppose that  $y$  is an MA( $\infty$ ) process satisfying Assumption 1. For a fixed  $h$  as  $T \rightarrow \infty$ , it holds that  $\sqrt{T}(\widehat{\gamma}_{h,J_T}^* - \widehat{\gamma}_h) = o_p(1)$ .

## 4 Finite Sample Performances

### 4.1 Univariate Analysis

In this section, we examine the finite sample performance of our circular-based autocovariance estimator in a univariate setting and compare it with the original  $\mathbf{A}$ -estimator of [Vogelsang and Yang \(2016\)](#) and our generalized  $\mathbf{A}$ -estimator. Following [Vogelsang and Yang \(2016\)](#), we assume that a univariate time series  $y$  follows an ARMA(1,1) process with unknown mean:

$$\begin{aligned} y_t &= \mu + u_t, \quad t \in [1, T], \\ u_t &= \phi u_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}, \\ \varepsilon_t &\sim i.i.d. \mathcal{N}(0, 1), \end{aligned} \tag{4.1}$$

where  $\mu$  is set to 0 without loss of generality. We initialize  $y_0 = u_0 = 0$  with a burn-in period of 100 observations to eliminate the impact of initial conditions. All simulation results are based on 100,000 replications.

We focus on the bias, variance and MSE of the elements in  $\widehat{\boldsymbol{\psi}}_h^* = \{\widehat{\psi}_{j,h}^*\}'_{j=0,h}$ , and compare each  $\widehat{\psi}_{j,h}^*$  with its  $\mathbf{A}$ -estimator and generalized  $\mathbf{A}$ -estimator counterparts, namely  $\widetilde{\gamma}_{j,h}^*$  and  $\widehat{\gamma}_{j,h,J_T}^*$ . Following the simulation design of [Vogelsang and Yang \(2016\)](#), we choose  $T \in \{25, 100\}$  with the parameter choices of  $\phi \in \{0, 0.5, 0.9\}$  and  $\theta \in \{0, 0.4, -0.4\}$ . We also include the circular and conventional sample autocovariance estimators ( $\widehat{\psi}_j$  and  $\widehat{\gamma}_j$ ) in the simulation as benchmarks. To save space, we present simulation results for  $j \in \{0, 1, 4, 8\}$  and  $h \in \{j, j+1, j+2\}$ . Results based on other values of  $h$  and  $j$  are qualitatively similar and are available upon request.

Table 1 shows the bias, variance and MSE results for sample variance estimators, namely  $\widehat{\psi}_0$ ,  $\widehat{\gamma}_0$ ,  $\widehat{\psi}_{0,h}^*$ ,  $\widetilde{\gamma}_{0,h}^*$  and  $\widehat{\gamma}_{0,h,J_T}^*$ . The following patterns can be observed from the table: (1) increasing  $h$  reduces the biases of all bias-corrected estimators but also unsurprisingly increases their variances and MSEs. MSE gains are only observed for the case with high persistence and small  $T$ , confirming the findings in [Vogelsang and Yang \(2016\)](#). (2) As  $T$  increases, both bias and variance of all estimators decreases. Importantly, the differences among  $\widehat{\psi}_{0,h}^*$ ,  $\widetilde{\gamma}_{0,h}^*$  and  $\widehat{\gamma}_{0,h,J_T}^*$  are also reduced. This demonstrates the consistency of the estimators and the asymptotic equivalence among the bias-corrected estimators as derived in Theorems 3 and 4; (3) For all MA(1) models (including MA(0)), all bias-corrected estimators are exactly unbiased, which corroborates our results in Theorem 2 and Proposition 5; (4) Among the three bias-corrected estimators,  $\widehat{\psi}_{0,h}^*$  performs slightly better than  $\widetilde{\gamma}_{0,h}^*$  in terms of variance and MSE with small  $T$  and high persistence, but its bias-reduction is also slightly worse than  $\widetilde{\gamma}_{0,h}^*$ . In most of the cases, the difference between the two estimators are indistinguishable.  $\widehat{\gamma}_{0,h,J_T}^*$  has relatively the largest bias among the three estimators, but it also has the smallest MSE for almost all cases. It is worth noting that  $\widehat{\gamma}_{0,0,J_T}^*$  does not reduce to  $\widetilde{\gamma}_{0,0}^*$  or  $\widehat{\psi}_{0,h}^*$  due to the extra sample moments used.

Results of  $\hat{\psi}_1$ ,  $\hat{\gamma}_1$ ,  $\hat{\psi}_{1,h}^*$ ,  $\tilde{\gamma}_{1,h}^*$  and  $\hat{\gamma}_{1,h,J_T}^*$  are reported in Table 2, which are largely consistent with the findings in Table 1. It is worth pointing out that for the case with  $T = 25$ , we can observe some slight differences between  $\hat{\psi}_1$  and  $\hat{\gamma}_1$ . In general,  $\hat{\psi}_1$  is more biased than  $\hat{\gamma}_1$  and has a slightly larger variance, corroborating our discussion in Remark 1. This discrepancy is more observable on higher-order autocovariance estimates, which we present in Tables 3 and 4. Overall, we find that  $\hat{\psi}_j$  is inferior to  $\hat{\gamma}_j$  with a larger bias and variance. The performances of bias-corrected estimators are again qualitatively similar to those in Tables 1 and 2, with more pronounced differences in terms of bias, variance and MSE. Interestingly, when the persistence is high ( $\phi = 0.9$ ),  $\hat{\psi}_{j,h}^*$  can have a considerably smaller variance than the  $\tilde{\gamma}_{j,h}^*$  which outweighs its larger bias, while  $\tilde{\gamma}_{j,h}^*$  is on average better than  $\hat{\psi}_{j,h}^*$  for the moderate to no persistence case, especially when the sample size is small.  $\hat{\gamma}_{j,h,J_T}^*$  consistently outperforms the other two bias-corrected estimators in terms of the MSE but also has the largest bias.

Tables 1 to 4 also provide some guidance on choosing  $h$  empirically. The findings clearly show that increasing  $h$  always leads to an inflated MSE while the bias reduction is in effect for relatively small  $h$ . In fact, we see nontrivial bias reduction from all three bias-corrected estimators for estimating  $\gamma_j$  even with  $h = j + 1$  unless a strong AR component is present. Therefore, empirically it is desirable to choose an  $h$  that covers the lags with the most prominent autocorrelation, but in general as small as possible to fully exploit the bias correction mechanism and avoid a largely inflated MSE. An adaptive choice of  $h$  remains an ongoing research question of this topic.

In summary, the simulation results above confirm that all three bias-corrected estimators ( $\hat{\psi}_{j,h}^*$ ,  $\tilde{\gamma}_{j,h}^*$  and  $\hat{\gamma}_{j,h,J_T}^*$ ) are exactly unbiased when the MA( $h$ ) assumption holds, and have significantly reduced bias under the ARMA(1,1) specifications considered relative to  $\hat{\psi}_j$  and  $\hat{\gamma}_j$ . For the MA(1) specifications, we find that  $\hat{\gamma}_{j,h,J_T}^*$  has the lowest MSE, followed by  $\tilde{\gamma}_{j,h}^*$  and  $\hat{\psi}_{j,h}^*$ , while the relative performances for the ARMA(1,1) models depend on the parameter choices. Overall, all three bias-corrected estimators have comparable finite sample performances with significant bias reduction even for small  $h$ , and some gains in MSE can be obtained when the data is highly persistent.

## 4.2 Multivariate Analysis

In this section, we conduct a multivariate simulation study to validate our theoretical results on bias-correction for cross-covariances. For simplicity, we simulate a VARMA(1,1) model as follows:

$$Y_t = \Phi_1 Y_{t-1} + U_t + \Theta_1 U_{t-1}, \quad U_t \sim i.i.d.\mathcal{N}(\mathbf{0}_{N \times 1}, \mathbf{\Omega}), \quad (4.2)$$

where  $\Phi_1$  and  $\Theta_1$  are  $N$ -by- $N$  coefficient matrices that satisfy the usual stationarity and invertibility condition, and  $\mathbf{\Omega}$  is a  $N$ -by- $N$  positive definite variance-covariance matrix. The true cross-covariances

Table 1: Bias, Variance and MSE of  $\hat{\psi}_0$ ,  $\hat{\gamma}_0$ ,  $\hat{\psi}_{0,h}^*$ ,  $\tilde{\gamma}_{0,h}^*$  and  $\hat{\gamma}_{0,h,J_T}^*$

$\theta$	$\phi$	$T = 25$	$\hat{\psi}_0$	$\hat{\gamma}_0$	$\hat{\psi}_{0,0}^*$	$\tilde{\gamma}_{0,0}^*$	$\hat{\gamma}_{0,0,J_T}^*$	$\hat{\psi}_{0,1}^*$	$\tilde{\gamma}_{0,1}^*$	$\hat{\gamma}_{0,1,J_T}^*$	$\hat{\psi}_{0,2}^*$	$\tilde{\gamma}_{0,2}^*$	$\hat{\gamma}_{0,2,J_T}^*$	
0	0	Bias	-0.040	-0.040	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
		Var	0.077	0.077	0.083	0.083	0.084	0.084	0.084	0.084	0.084	0.084	0.084	
		MSE	0.079	0.079	0.083	0.083	0.084	0.084	0.084	0.084	0.084	0.084	0.084	
	0.5	Bias	-0.151	-0.151	-0.102	-0.102	-0.115	-0.053	-0.053	-0.053	-0.060	-0.028	-0.028	-0.032
		Var	0.185	0.185	0.201	0.201	0.192	0.228	0.228	0.228	0.222	0.250	0.250	0.246
		MSE	0.208	0.208	0.211	0.211	0.205	0.231	0.231	0.231	0.225	0.251	0.251	0.247
	0.9	Bias	-2.591	-2.591	-2.480	-2.480	-2.600	-2.290	-2.287	-2.400	-2.123	-2.114	-2.220	
		Var	3.323	3.323	3.606	3.606	3.088	4.215	4.234	3.660	4.882	4.951	4.316	
		MSE	10.036	10.036	9.754	9.754	9.848	9.459	9.466	9.418	9.391	9.421	9.244	
0.4	0	Bias	-0.077	-0.077	-0.032	-0.032	-0.037	0.000	0.000	0.000	0.000	0.000	0.000	
		Var	0.121	0.121	0.131	0.131	0.129	0.144	0.144	0.143	0.147	0.147	0.147	
		MSE	0.127	0.127	0.132	0.132	0.130	0.144	0.144	0.143	0.147	0.147	0.147	
	0.5	Bias	-0.295	-0.295	-0.221	-0.221	-0.249	-0.115	-0.115	-0.130	-0.060	-0.060	-0.068	
		Var	0.584	0.584	0.634	0.634	0.600	0.734	0.734	0.707	0.819	0.818	0.799	
		MSE	0.671	0.671	0.683	0.683	0.662	0.747	0.747	0.724	0.822	0.822	0.804	
	0.9	Bias	-5.072	-5.072	-4.871	-4.871	-5.107	-4.499	-4.493	-4.713	-4.171	-4.153	-4.361	
		Var	12.467	12.467	13.527	13.527	11.562	15.820	15.894	13.714	18.343	18.606	16.187	
		MSE	38.188	38.188	37.251	37.251	37.646	36.057	36.080	35.929	35.740	35.850	35.202	
-0.4	0	Bias	-0.016	-0.016	-0.001	-0.001	-0.001	0.000	0.000	0.000	-0.001	-0.001	-0.001	
		Var	0.132	0.132	0.131	0.131	0.131	0.133	0.133	0.133	0.133	0.133	0.133	
		MSE	0.132	0.132	0.131	0.131	0.131	0.133	0.133	0.133	0.133	0.133	0.133	
	0.5	Bias	-0.056	-0.056	-0.009	-0.009	-0.010	-0.004	-0.004	-0.005	-0.003	-0.003	-0.003	
		Var	0.078	0.078	0.088	0.088	0.088	0.090	0.090	0.090	0.092	0.091	0.091	
		MSE	0.081	0.081	0.088	0.088	0.088	0.090	0.090	0.090	0.092	0.092	0.091	
	0.9	Bias	-0.939	-0.939	-0.814	-0.814	-0.853	-0.755	-0.752	-0.789	-0.703	-0.696	-0.732	
		Var	0.503	0.503	0.633	0.636	0.557	0.727	0.737	0.650	0.827	0.848	0.753	
		MSE	1.385	1.385	1.296	1.297	1.285	1.298	1.302	1.274	1.322	1.333	1.289	
0	0	Bias	-0.010	-0.010	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
		Var	0.020	0.020	0.020	0.020	0.020	0.020	0.020	0.020	0.020	0.020	0.020	
		MSE	0.020	0.020	0.020	0.020	0.020	0.020	0.020	0.020	0.020	0.020	0.020	
	0.5	Bias	-0.040	-0.040	-0.027	-0.027	-0.031	-0.014	-0.014	-0.016	-0.007	-0.007	-0.008	
		Var	0.056	0.056	0.057	0.057	0.056	0.059	0.059	0.058	0.060	0.060	0.060	
		MSE	0.057	0.057	0.057	0.057	0.057	0.059	0.059	0.058	0.060	0.060	0.060	
	0.9	Bias	-0.909	-0.909	-0.865	-0.865	-0.963	-0.786	-0.786	-0.875	-0.714	-0.714	-0.796	
		Var	3.337	3.337	3.405	3.405	3.120	3.545	3.545	3.263	3.688	3.689	3.411	
		MSE	4.163	4.163	4.153	4.153	4.047	4.162	4.162	4.029	4.199	4.199	4.045	
	0.4	0	Bias	-0.020	-0.020	-0.009	-0.009	-0.010	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001
			Var	0.032	0.032	0.033	0.033	0.033	0.034	0.034	0.034	0.034	0.034	0.034
			MSE	0.033	0.033	0.033	0.033	0.033	0.034	0.034	0.034	0.034	0.034	0.034
		0.5	Bias	-0.078	-0.078	-0.058	-0.058	-0.067	-0.030	-0.030	-0.034	-0.015	-0.015	-0.018
			Var	0.183	0.183	0.186	0.186	0.184	0.193	0.193	0.191	0.198	0.198	0.197
			MSE	0.189	0.189	0.190	0.190	0.188	0.194	0.194	0.192	0.199	0.199	0.198
		0.9	Bias	-1.780	-1.780	-1.698	-1.698	-1.891	-1.543	-1.543	-1.719	-1.403	-1.403	-1.564
			Var	12.732	12.732	12.990	12.990	11.899	13.523	13.523	12.444	14.073	14.074	13.010
			MSE	15.901	15.901	15.875	15.875	15.476	15.905	15.905	15.401	16.042	16.042	15.455
-0.4		0	Bias	-0.004	-0.004	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001
			Var	0.034	0.034	0.033	0.033	0.033	0.034	0.034	0.034	0.034	0.034	0.034
			MSE	0.034	0.034	0.033	0.033	0.033	0.034	0.034	0.034	0.034	0.034	0.034
		0.5	Bias	-0.015	-0.015	-0.003	-0.003	-0.003	-0.002	-0.002	-0.002	-0.001	-0.001	-0.001
			Var	0.021	0.021	0.021	0.021	0.021	0.022	0.022	0.022	0.022	0.022	0.022
			MSE	0.021	0.021	0.021	0.021	0.021	0.022	0.022	0.022	0.022	0.022	0.022
		0.9	Bias	-0.328	-0.328	-0.280	-0.280	-0.311	-0.254	-0.254	-0.283	-0.231	-0.231	-0.258
			Var	0.453	0.453	0.481	0.481	0.444	0.500	0.500	0.464	0.520	0.520	0.484
			MSE	0.561	0.561	0.559	0.559	0.541	0.565	0.565	0.544	0.573	0.573	0.551

Note: the data generating process is specified in Eq. (4.1). MSE stands for mean squared error. All results are generated based on 100,000 Monte Carlo draws.



Table 2: Bias, Variance and MSE of  $\hat{\psi}_1$ ,  $\hat{\gamma}_1$ ,  $\hat{\psi}_{1,h}^*$ ,  $\hat{\gamma}_{1,h}^*$  and  $\hat{\gamma}_{1,h,J_T}^*$

$\theta$	$\phi$	$T = 25$	$\hat{\psi}_1$	$\hat{\gamma}_1$	$\hat{\psi}_{1,1}^*$	$\hat{\gamma}_{1,1}^*$	$\hat{\gamma}_{1,1,J_T}^*$	$\hat{\psi}_{1,2}^*$	$\hat{\gamma}_{1,2}^*$	$\hat{\gamma}_{1,2,J_T}^*$	$\hat{\psi}_{1,3}^*$	$\hat{\gamma}_{1,3}^*$	$\hat{\gamma}_{1,3,J_T}^*$	
0	0	Bias	-0.039	-0.038	0.001	0.000	0.000	0.001	0.001	0.001	0.000	0.000	0.000	
		Var	0.037	0.035	0.047	0.046	0.046	0.048	0.046	0.046	0.046	0.048	0.046	0.046
		MSE	0.038	0.037	0.047	0.046	0.046	0.048	0.046	0.046	0.046	0.048	0.046	0.046
	0.5	Bias	-0.178	-0.175	-0.055	-0.055	-0.069	-0.029	-0.028	-0.028	-0.035	-0.015	-0.015	-0.018
		Var	0.127	0.125	0.178	0.175	0.162	0.203	0.199	0.199	0.190	0.223	0.218	0.213
		MSE	0.158	0.155	0.182	0.178	0.167	0.204	0.200	0.192	0.192	0.223	0.218	0.213
	0.9	Bias	-2.764	-2.728	-2.369	-2.333	-2.559	-2.195	-2.154	-2.349	-2.043	-1.992	-1.992	-2.164
		Var	2.679	2.835	3.784	3.980	2.908	4.460	4.710	3.607	5.189	5.532	4.393	
		MSE	10.319	10.276	9.394	9.424	9.457	9.277	9.348	9.126	9.363	9.498	9.076	
0.4	0	Bias	-0.092	-0.091	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
		Var	0.060	0.058	0.084	0.081	0.079	0.092	0.089	0.089	0.089	0.095	0.091	0.092
		MSE	0.068	0.066	0.084	0.081	0.079	0.092	0.089	0.089	0.089	0.095	0.091	0.092
	0.5	Bias	-0.353	-0.348	-0.120	-0.119	-0.149	-0.063	-0.062	-0.077	-0.033	-0.033	-0.040	
		Var	0.431	0.426	0.609	0.601	0.549	0.701	0.690	0.655	0.778	0.764	0.742	
		MSE	0.555	0.548	0.624	0.615	0.571	0.705	0.694	0.661	0.780	0.765	0.744	
	0.9	Bias	-5.412	-5.342	-4.653	-4.583	-5.027	-4.312	-4.231	-4.615	-4.013	-3.912	-4.251	
		Var	10.066	10.672	14.221	14.986	10.910	16.771	17.745	13.545	19.527	20.857	16.515	
		MSE	39.359	39.206	35.872	35.993	36.182	35.363	35.643	34.843	35.635	36.162	34.588	
-0.4	0	Bias	0.001	0.002	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	
		Var	0.071	0.069	0.085	0.082	0.084	0.078	0.076	0.075	0.079	0.077	0.077	
		MSE	0.071	0.069	0.085	0.082	0.084	0.078	0.076	0.075	0.079	0.077	0.077	
	0.5	Bias	-0.060	-0.058	-0.008	-0.008	-0.011	-0.004	-0.004	-0.005	-0.002	-0.002	-0.003	
		Var	0.041	0.039	0.054	0.052	0.051	0.056	0.054	0.054	0.054	0.058	0.056	
		MSE	0.044	0.042	0.054	0.052	0.051	0.056	0.054	0.054	0.054	0.058	0.056	
	0.9	Bias	-1.000	-0.987	-0.842	-0.829	-0.910	-0.780	-0.766	-0.835	-0.726	-0.708	-0.769	
		Var	0.397	0.415	0.559	0.582	0.435	0.655	0.685	0.535	0.758	0.799	0.646	
		MSE	1.397	1.389	1.268	1.270	1.263	1.264	1.271	1.232	1.285	1.301	1.238	
0.4	0	Bias	-0.010	-0.010	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
		Var	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	
		MSE	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	
	0.5	Bias	-0.046	-0.046	-0.014	-0.014	-0.018	-0.007	-0.007	-0.009	-0.004	-0.004	-0.005	
		Var	0.042	0.042	0.046	0.045	0.044	0.047	0.047	0.046	0.048	0.048	0.047	
		MSE	0.044	0.044	0.046	0.046	0.045	0.047	0.047	0.046	0.048	0.048	0.047	
	0.9	Bias	-0.956	-0.954	-0.794	-0.793	-0.973	-0.721	-0.720	-0.881	-0.656	-0.655	-0.799	
		Var	3.237	3.238	3.511	3.511	2.971	3.656	3.657	3.135	3.804	3.805	3.303	
		MSE	4.150	4.149	4.140	4.140	3.917	4.176	4.175	3.912	4.235	4.234	3.941	
	0.4	0	Bias	-0.024	-0.023	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
			Var	0.017	0.017	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019
			MSE	0.018	0.018	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019
		0.5	Bias	-0.092	-0.092	-0.029	-0.029	-0.039	-0.015	-0.015	-0.020	-0.008	-0.008	-0.010
			Var	0.148	0.147	0.160	0.160	0.156	0.166	0.165	0.163	0.170	0.169	0.168
			MSE	0.156	0.156	0.161	0.161	0.157	0.166	0.165	0.163	0.170	0.169	0.168
		0.9	Bias	-1.873	-1.871	-1.559	-1.557	-1.911	-1.417	-1.415	-1.731	-1.289	-1.287	-1.570
			Var	12.361	12.366	13.408	13.411	11.339	13.962	13.966	11.964	14.530	14.533	12.605
			MSE	15.870	15.865	15.838	15.835	14.989	15.971	15.969	14.961	16.191	16.189	15.069
-0.4	0	Bias	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
		Var	0.018	0.018	0.019	0.019	0.019	0.019	0.018	0.018	0.018	0.019	0.019	
		MSE	0.018	0.018	0.019	0.019	0.019	0.019	0.018	0.018	0.018	0.019	0.019	
	0.5	Bias	-0.015	-0.015	-0.002	-0.002	-0.003	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	
		Var	0.011	0.011	0.012	0.012	0.012	0.012	0.012	0.012	0.012	0.012	0.012	
		MSE	0.012	0.011	0.012	0.012	0.012	0.012	0.012	0.012	0.012	0.012	0.012	
	0.9	Bias	-0.344	-0.344	-0.282	-0.282	-0.346	-0.256	-0.256	-0.313	-0.233	-0.233	-0.284	
		Var	0.435	0.435	0.472	0.472	0.401	0.491	0.491	0.423	0.511	0.511	0.445	
		MSE	0.554	0.553	0.551	0.551	0.521	0.557	0.557	0.521	0.565	0.565	0.526	

Note: the data generating process is specified in Eq. (4.1). MSE stands for mean squared error. All results are generated based on 100,000 Monte Carlo draws.

Table 3: Bias, Variance and MSE of  $\hat{\psi}_4$ ,  $\hat{\gamma}_4$ ,  $\hat{\psi}_{4,h}^*$ ,  $\hat{\gamma}_{4,h}^*$  and  $\hat{\gamma}_{4,h,J_T}^*$ 

$\theta$	$\phi$	$T = 25$	$\hat{\psi}_4$	$\hat{\gamma}_4$	$\hat{\psi}_{4,4}^*$	$\hat{\gamma}_{4,4}^*$	$\hat{\gamma}_{4,4,J_T}^*$	$\hat{\psi}_{4,5}^*$	$\hat{\gamma}_{4,5}^*$	$\hat{\gamma}_{4,5,J_T}^*$	$\hat{\psi}_{4,6}^*$	$\hat{\gamma}_{4,6}^*$	$\hat{\gamma}_{4,6,J_T}^*$		
0	0	Bias	-0.040	-0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000		
		Var	0.037	0.031	0.064	0.053	0.054	0.065	0.054	0.054	0.066	0.055	0.055		
		MSE	0.038	0.032	0.064	0.053	0.054	0.065	0.054	0.054	0.066	0.055	0.055		
	0.5	Bias	-0.166	-0.147	-0.011	-0.010	-0.012	-0.007	-0.006	-0.007	-0.007	-0.005	-0.004	-0.005	
		Var	0.069	0.060	0.169	0.144	0.131	0.204	0.170	0.163	0.163	0.235	0.193	0.189	
		MSE	0.097	0.081	0.169	0.144	0.131	0.204	0.170	0.163	0.235	0.193	0.189		
	0.9	Bias	-3.055	-2.859	-2.099	-1.919	-2.117	-1.970	-1.773	-1.951	-1.951	-1.861	-1.640	-1.803	
		Var	0.724	1.084	3.276	4.154	2.882	4.033	5.098	3.751	4.828	6.158	4.726		
		MSE	10.059	9.256	7.681	7.836	7.364	7.913	8.241	7.558	8.290	8.847	7.976		
0.4	0	Bias	-0.078	-0.066	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	
		Var	0.054	0.046	0.103	0.086	0.084	0.119	0.098	0.099	0.099	0.126	0.103	0.104	
		MSE	0.060	0.050	0.103	0.086	0.084	0.119	0.098	0.099	0.126	0.103	0.104		
	0.5	Bias	-0.327	-0.290	-0.024	-0.022	-0.027	-0.016	-0.014	-0.016	-0.016	-0.011	-0.010	-0.011	
		Var	0.204	0.178	0.543	0.465	0.415	0.668	0.562	0.529	0.786	0.648	0.630		
		MSE	0.310	0.262	0.543	0.466	0.416	0.668	0.562	0.530	0.786	0.648	0.631		
	0.9	Bias	-5.983	-5.600	-4.122	-3.769	-4.158	-3.868	-3.482	-3.832	-3.653	-3.220	-3.541		
		Var	2.673	4.073	12.282	15.694	10.848	15.129	19.268	14.132	18.124	23.293	17.825		
		MSE	38.472	35.433	29.270	29.896	28.138	30.089	31.394	28.815	31.469	33.664	30.360		
-0.4	0	Bias	-0.015	-0.011	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	
		Var	0.065	0.055	0.105	0.088	0.091	0.092	0.078	0.078	0.078	0.095	0.080	0.081	
		MSE	0.065	0.055	0.105	0.088	0.091	0.092	0.078	0.078	0.078	0.095	0.080	0.081	
	0.5	Bias	-0.058	-0.050	-0.002	-0.001	-0.002	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	
		Var	0.036	0.030	0.066	0.055	0.055	0.071	0.059	0.059	0.059	0.075	0.061	0.061	
		MSE	0.039	0.033	0.066	0.055	0.055	0.071	0.059	0.059	0.075	0.061	0.061		
	0.9	Bias	-1.104	-1.031	-0.747	-0.683	-0.753	-0.701	-0.631	-0.694	-0.662	-0.584	-0.641		
		Var	0.122	0.165	0.496	0.600	0.427	0.606	0.732	0.551	0.721	0.879	0.687		
		MSE	1.342	1.229	1.054	1.066	0.994	1.097	1.130	1.032	1.159	1.219	1.099		
0.4	0	Bias	-0.010	-0.009	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
		Var	0.010	0.009	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011	
		MSE	0.010	0.010	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011	
	0.5	Bias	-0.043	-0.042	-0.002	-0.002	-0.003	-0.001	-0.001	-0.002	-0.002	-0.001	-0.001	-0.001	
		Var	0.027	0.026	0.033	0.032	0.031	0.034	0.033	0.033	0.033	0.035	0.034	0.034	
		MSE	0.029	0.028	0.033	0.032	0.031	0.034	0.033	0.033	0.033	0.035	0.034	0.034	
	0.9	Bias	-1.047	-1.036	-0.616	-0.611	-0.751	-0.561	-0.556	-0.682	-0.511	-0.506	-0.619		
		Var	2.653	2.644	3.507	3.483	2.962	3.659	3.634	3.133	3.813	3.786	3.306		
		MSE	3.749	3.717	3.886	3.857	3.527	3.973	3.943	3.598	4.074	4.042	3.689		
	0.4	0	Bias	-0.020	-0.019	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
			Var	0.016	0.015	0.018	0.018	0.018	0.019	0.018	0.018	0.018	0.019	0.018	0.018
			MSE	0.016	0.016	0.018	0.018	0.018	0.019	0.018	0.018	0.018	0.019	0.018	0.018
		0.5	Bias	-0.085	-0.083	-0.005	-0.005	-0.006	-0.003	-0.003	-0.003	-0.003	-0.002	-0.002	-0.002
			Var	0.088	0.085	0.110	0.107	0.103	0.114	0.111	0.109	0.118	0.114	0.113	
			MSE	0.095	0.092	0.110	0.107	0.103	0.114	0.111	0.109	0.118	0.114	0.113	
		0.9	Bias	-2.052	-2.031	-1.210	-1.201	-1.476	-1.101	-1.092	-1.339	-1.003	-0.994	-1.215	
			Var	10.144	10.110	13.409	13.321	11.322	13.992	13.897	11.976	14.583	14.479	12.639	
			MSE	14.355	14.234	14.872	14.763	13.502	15.205	15.090	13.770	15.588	15.468	14.116	
-0.4	0	Bias	-0.003	-0.003	0.001	0.001	0.001	0.001	0.001	0.000	0.001	0.001	0.000	0.000	
		Var	0.017	0.016	0.018	0.018	0.018	0.018	0.017	0.017	0.017	0.018	0.017	0.017	
		MSE	0.017	0.016	0.018	0.018	0.018	0.018	0.017	0.017	0.017	0.018	0.017	0.017	
	0.5	Bias	-0.015	-0.014	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
		Var	0.010	0.010	0.012	0.011	0.011	0.012	0.011	0.011	0.011	0.012	0.011	0.011	
		MSE	0.010	0.010	0.012	0.011	0.011	0.012	0.011	0.011	0.011	0.012	0.011	0.011	
	0.9	Bias	-0.377	-0.373	-0.219	-0.217	-0.267	-0.199	-0.198	-0.242	-0.181	-0.180	-0.220		
		Var	0.355	0.353	0.468	0.465	0.397	0.488	0.484	0.419	0.509	0.504	0.442		
		MSE	0.497	0.492	0.516	0.512	0.468	0.528	0.523	0.478	0.542	0.537	0.491		

Note: the data generating process is specified in Eq. (4.1). MSE stands for mean squared error. All results are generated based on 100,000 Monte Carlo draws.

Table 4: Bias, Variance and MSE of  $\hat{\psi}_8$ ,  $\hat{\gamma}_8$ ,  $\hat{\psi}_{8,h}^*$ ,  $\hat{\gamma}_{8,h}^*$  and  $\hat{\gamma}_{8,h,J_T}^*$

$\theta$	$\phi$	$T = 25$	$\hat{\psi}_8$	$\hat{\gamma}_8$	$\hat{\psi}_{8,8}^*$	$\hat{\gamma}_{8,8}^*$	$\hat{\gamma}_{8,8,J_T}^*$	$\hat{\psi}_{8,9}^*$	$\hat{\gamma}_{8,9}^*$	$\hat{\gamma}_{8,9,J_T}^*$	$\hat{\psi}_{8,10}^*$	$\hat{\gamma}_{8,10}^*$	$\hat{\gamma}_{8,10,J_T}^*$	
0	0	Bias	-0.039	-0.027	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	
		Var	0.037	0.025	0.109	0.069	0.070	0.116	0.070	0.072	0.131	0.072	0.074	
		MSE	0.038	0.026	0.109	0.069	0.070	0.116	0.070	0.072	0.131	0.072	0.074	
	0.5	Bias	-0.152	-0.107	-0.002	-0.001	-0.002	-0.002	-0.001	-0.001	-0.002	-0.002	-0.001	-0.002
		Var	0.069	0.052	0.239	0.176	0.152	0.336	0.223	0.206	0.439	0.266	0.254	
		MSE	0.092	0.064	0.239	0.176	0.152	0.336	0.223	0.206	0.439	0.266	0.254	
	0.9	Bias	-3.037	-2.594	-1.817	-1.413	-1.581	-1.748	-1.300	-1.456	-1.700	-1.195	-1.339	
		Var	0.593	0.277	1.459	3.783	2.238	2.037	4.993	3.236	2.584	6.430	4.449	
		MSE	9.819	7.004	4.760	5.780	4.736	5.093	6.684	5.356	5.475	7.856	6.242	
0.4	0	Bias	-0.076	-0.052	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	
		Var	0.054	0.038	0.173	0.112	0.107	0.232	0.133	0.136	0.280	0.143	0.148	
		MSE	0.060	0.041	0.173	0.112	0.107	0.232	0.133	0.136	0.280	0.143	0.148	
	0.5	Bias	-0.297	-0.210	-0.004	-0.003	-0.004	-0.004	-0.003	-0.003	-0.003	-0.004	-0.003	-0.003
		Var	0.202	0.158	0.724	0.561	0.464	1.054	0.728	0.655	1.404	0.888	0.834	
		MSE	0.291	0.202	0.724	0.561	0.464	1.054	0.728	0.655	1.404	0.888	0.834	
	0.9	Bias	-5.949	-5.083	-3.568	-2.775	-3.104	-3.433	-2.553	-2.859	-3.338	-2.345	-2.629	
		Var	2.178	0.994	5.263	14.233	8.369	7.353	18.796	12.121	9.323	24.228	16.696	
		MSE	37.571	26.828	17.995	21.933	18.004	19.140	25.316	20.295	20.468	29.727	23.605	
-0.4	0	Bias	-0.015	-0.010	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	
		Var	0.065	0.044	0.181	0.113	0.120	0.144	0.097	0.097	0.162	0.101	0.103	
		MSE	0.065	0.044	0.181	0.113	0.120	0.144	0.097	0.097	0.162	0.101	0.103	
	0.5	Bias	-0.056	-0.039	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	-0.001	
		Var	0.036	0.025	0.110	0.071	0.070	0.127	0.077	0.077	0.149	0.083	0.084	
		MSE	0.039	0.027	0.110	0.071	0.070	0.127	0.077	0.077	0.149	0.083	0.084	
	0.9	Bias	-1.097	-0.934	-0.646	-0.503	-0.562	-0.622	-0.463	-0.518	-0.605	-0.425	-0.477	
		Var	0.104	0.055	0.277	0.562	0.349	0.377	0.734	0.494	0.476	0.937	0.668	
		MSE	1.308	0.927	0.695	0.814	0.665	0.764	0.948	0.763	0.842	1.118	0.895	
0	0	Bias	-0.010	-0.010	0.000	0.000	-0.001	0.000	0.000	-0.001	0.000	0.000	0.000	
		Var	0.010	0.009	0.012	0.011	0.011	0.012	0.011	0.011	0.011	0.012	0.011	
		MSE	0.010	0.009	0.012	0.011	0.011	0.012	0.011	0.011	0.011	0.012	0.011	
	0.5	Bias	-0.040	-0.038	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	
		Var	0.026	0.024	0.035	0.033	0.032	0.037	0.034	0.033	0.038	0.035	0.034	
		MSE	0.028	0.026	0.035	0.033	0.032	0.037	0.034	0.033	0.038	0.035	0.034	
	0.9	Bias	-1.089	-1.056	-0.441	-0.432	-0.532	-0.403	-0.393	-0.484	-0.367	-0.358	-0.439	
		Var	1.860	1.824	3.289	3.188	2.695	3.446	3.337	2.864	3.604	3.487	3.035	
		MSE	3.047	2.939	3.484	3.374	2.979	3.608	3.491	3.098	3.739	3.615	3.229	
	0.4	0	Bias	-0.020	-0.019	0.000	-0.001	-0.001	0.000	-0.001	-0.001	-0.001	0.000	-0.001
			Var	0.016	0.015	0.020	0.018	0.018	0.020	0.019	0.019	0.019	0.021	0.019
			MSE	0.016	0.015	0.020	0.018	0.018	0.020	0.019	0.019	0.019	0.021	0.019
		0.5	Bias	-0.079	-0.074	-0.001	-0.001	-0.002	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001
			Var	0.085	0.079	0.117	0.108	0.105	0.122	0.113	0.111	0.125	0.116	0.115
			MSE	0.091	0.084	0.117	0.108	0.105	0.122	0.113	0.111	0.125	0.116	0.115
		0.9	Bias	-2.135	-2.071	-0.867	-0.848	-1.046	-0.790	-0.772	-0.950	-0.721	-0.704	-0.863
			Var	7.107	6.968	12.577	12.193	10.303	13.176	12.763	10.949	13.785	13.339	11.604
			MSE	11.664	11.256	13.329	12.913	11.396	13.801	13.360	11.851	14.305	13.834	12.349
-0.4	0	Bias	-0.004	-0.004	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
		Var	0.017	0.015	0.020	0.018	0.019	0.020	0.018	0.018	0.020	0.018	0.018	
		MSE	0.017	0.015	0.020	0.018	0.019	0.020	0.018	0.018	0.020	0.018	0.018	
	0.5	Bias	-0.015	-0.014	0.000	-0.001	-0.001	0.000	-0.001	-0.001	0.000	-0.001	-0.001	
		Var	0.010	0.009	0.013	0.012	0.012	0.013	0.012	0.012	0.013	0.012	0.012	
		MSE	0.010	0.009	0.013	0.012	0.012	0.013	0.012	0.012	0.013	0.012	0.012	
	0.9	Bias	-0.392	-0.380	-0.157	-0.154	-0.190	-0.143	-0.140	-0.172	-0.131	-0.128	-0.156	
		Var	0.251	0.245	0.439	0.425	0.361	0.460	0.445	0.383	0.481	0.465	0.406	
		MSE	0.404	0.389	0.464	0.449	0.397	0.481	0.464	0.413	0.498	0.481	0.430	

Note: the data generating process is specified in Eq. (4.1). MSE stands for mean squared error. All results are generated based on 100,000 Monte Carlo draws.

of  $Y_t$  can be found via the following iterative relationship:

$$\Upsilon_0 = \mathbf{I}_N, \Upsilon_1 = \Phi_1 + \Theta_1, \Upsilon_j = \Phi_1 \Upsilon_{j-1}, \quad j > 1, \quad (4.3)$$

$$\Gamma_j = \sum_{k=0}^{\infty} \Upsilon_{j+k} \Omega \Upsilon_k'. \quad (4.4)$$

Following the simulation setting in [Koreisha and Pukkila \(1988, 1989\)](#); [Frutos and Serrano \(2002\)](#); [Kascha \(2012\)](#), we simulate two VARMA(1,1) models with  $N \in \{3, 5\}$ . The parameter setting for the  $N = 3$  case is:

$$\Phi_1 = \begin{bmatrix} 0.7, & 0, & 0 \\ 0, & 0, & 0 \\ 0, & 0.4, & 0 \end{bmatrix}, \quad \Theta_1 = \begin{bmatrix} 0, & 1.1, & 0 \\ 0, & -0.6, & 0 \\ 0, & 0, & -0.5 \end{bmatrix}, \quad \Omega = \begin{bmatrix} 1, & & & \\ -0.7, & 1 & & \\ -0.4, & 0, & 1 & \end{bmatrix}. \quad (4.5)$$

For the  $N = 5$  case, we use:

$$\Phi_1 = \begin{bmatrix} 0.5, & 0, & 0, & 0, & 0 \\ 0, & 0, & 0.8, & 0, & 0 \\ 0, & -0.4, & 0, & 0, & 0 \\ 0, & 0, & 0, & 0, & 0 \\ 0.2, & 0, & 0, & 0, & 0 \end{bmatrix}, \quad \Theta_1 = \begin{bmatrix} 0, & 0, & 0, & 1.1, & 0 \\ 0, & 0, & 0, & 0, & 0.2 \\ 0, & 0, & 0, & 0, & 0 \\ -0.55, & 0, & 0, & 0.8, & 0 \\ 0, & 0, & 0, & 0, & 0.6 \end{bmatrix}, \quad \Omega = \begin{bmatrix} 1 & & & & \\ 0.2, & 1 & & & \\ 0, & 0, & 1 & & \\ 0, & 0, & 0.7, & 1 & \\ 0, & 0, & 0, & -0.4, & 1 \end{bmatrix}. \quad (4.6)$$

According to [Koreisha and Pukkila \(1989\)](#); [Kascha \(2012\)](#), the above two specifications have low density of non-zero elements, broad variation in the magnitude of parameter variations and complex causal mechanisms which are typical for real data applications. We also consider the corresponding VMA(1) or VAR(1) model by setting  $\Phi_1 = \mathbf{0}_{N \times N}$  or  $\Theta_1 = \mathbf{0}_{N \times N}$  to further validate our results. Similar to the univariate analysis, we simulate the models with 100,000 replications and  $T \in \{25, 100\}$  for the four specifications considered with a burn-in period of 100 observations, and present the results for  $j \in \{0, 1, 4, 8\}$  with  $h \in \{j, j+1, j+2, j+3\}$ .

We firstly compare our estimator of cross-covariance matrices  $\{\hat{\Psi}_{j,h}^*\}_{j=0:h}$  with the simple circular and conventional cross-covariance estimators, namely  $\hat{\Psi}_j$  and  $\hat{\Gamma}_j$ . Each  $\hat{\Psi}_{j,h}^*$  can be easily computed by reversing the vectorization operation for the corresponding row of  $\hat{\Psi}_h^*$ . We summarize the overall performance of each estimator by calculating averages of the element-wise absolute bias, variance and mean squared error. Taking the estimator  $\hat{\Gamma}_j$  as an example, we compute the Average Bias (AB), Average Variance (AV) and Average Mean Squared Error (AMSE) as follows:

$$\begin{aligned} \text{AB}(\hat{\Gamma}_j) &= \frac{\iota'_{N \times 1} \text{E}[\Gamma_j - \hat{\Gamma}_j] \iota_{N \times 1}}{N^2}, \\ \text{AV}(\hat{\Gamma}_j) &= \frac{\iota'_{N \times 1} (\text{E}[\hat{\Gamma}_j \circ \hat{\Gamma}_j] - \text{E}[\hat{\Gamma}_j] \circ \text{E}[\hat{\Gamma}_j]) \iota_{N \times 1}}{N^2}, \\ \text{AMSE}(\hat{\Gamma}_j) &= \frac{\iota'_{N \times 1} \text{E}[(\Gamma_j - \hat{\Gamma}_j) \circ (\Gamma_j - \hat{\Gamma}_j)] \iota_{N \times 1}}{N^2}, \end{aligned} \quad (4.7)$$

where  $\circ$  denotes the Hadamard product, and  $\iota_{m \times n}$  is a  $m$ -by- $n$  matrix of ones. We present our simulation results in [Tables 5 and 6](#).

Table 5: Average bias, variance and MSE of  $\hat{\Psi}_{j,h}^*$ ,  $\hat{\Psi}_j$  and  $\hat{\Gamma}_j$  for  $N = 3$

Model	VMA(1)	$T = 25$						$T = 100$					
$j =$		$\hat{\Psi}_j$	$\hat{\Gamma}_j$	$\hat{\Psi}_{j,j}^*$	$\hat{\Psi}_{j,j+1}^*$	$\hat{\Psi}_{j,j+2}^*$	$\hat{\Psi}_{j,j+3}^*$	$\hat{\Psi}_j$	$\hat{\Gamma}_j$	$\hat{\Psi}_{j,j}^*$	$\hat{\Psi}_{j,j+1}^*$	$\hat{\Psi}_{j,j+2}^*$	$\hat{\Psi}_{j,j+3}^*$
0	AB	-0.004	-0.004	0.001	0.000	0.000	0.000	-0.001	-0.001	0.000	0.000	0.000	0.000
	AV	0.206	0.206	0.223	0.203	0.207	0.207	0.052	0.052	0.053	0.052	0.052	0.052
	AMSE	0.206	0.206	0.225	0.203	0.207	0.207	0.052	0.052	0.053	0.052	0.052	0.052
1	AB	-0.004	-0.004	0.000	0.000	0.000	0.000	-0.001	-0.001	0.000	0.000	0.000	0.000
	AV	0.143	0.138	0.166	0.153	0.156	0.157	0.036	0.036	0.038	0.037	0.037	0.037
	AMSE	0.143	0.139	0.166	0.153	0.156	0.157	0.036	0.036	0.038	0.037	0.037	0.037
4	AB	-0.004	-0.004	0.000	0.000	0.000	0.000	-0.001	-0.001	0.000	0.000	0.000	0.000
	AV	0.131	0.111	0.205	0.184	0.189	0.190	0.034	0.032	0.037	0.036	0.036	0.036
	AMSE	0.131	0.111	0.205	0.184	0.189	0.190	0.034	0.032	0.037	0.036	0.036	0.036
8	AB	-0.004	-0.003	0.000	0.000	0.000	0.000	-0.001	-0.001	0.000	0.000	0.000	0.000
	AV	0.131	0.090	0.345	0.280	0.310	0.373	0.033	0.031	0.040	0.039	0.039	0.039
	AMSE	0.132	0.090	0.345	0.280	0.310	0.373	0.033	0.031	0.040	0.039	0.039	0.039
Model	VAR(1)	$T = 25$						$T = 100$					
$j =$		$\hat{\Psi}_j$	$\hat{\Gamma}_j$	$\hat{\Psi}_{j,j}^*$	$\hat{\Psi}_{j,j+1}^*$	$\hat{\Psi}_{j,j+2}^*$	$\hat{\Psi}_{j,j+3}^*$	$\hat{\Psi}_j$	$\hat{\Gamma}_j$	$\hat{\Psi}_{j,j}^*$	$\hat{\Psi}_{j,j+1}^*$	$\hat{\Psi}_{j,j+2}^*$	$\hat{\Psi}_{j,j+3}^*$
0	AB	-0.020	-0.020	-0.013	-0.009	-0.006	-0.005	-0.005	-0.005	-0.004	-0.002	-0.002	-0.001
	AV	0.126	0.126	0.136	0.152	0.166	0.180	0.040	0.040	0.041	0.042	0.043	0.044
	AMSE	0.146	0.146	0.150	0.158	0.170	0.182	0.042	0.042	0.042	0.043	0.044	0.044
1	AB	-0.022	-0.022	-0.009	-0.007	-0.005	-0.004	-0.006	-0.006	-0.002	-0.002	-0.001	-0.001
	AV	0.101	0.098	0.135	0.151	0.166	0.181	0.035	0.035	0.038	0.039	0.039	0.040
	AMSE	0.127	0.124	0.142	0.155	0.168	0.182	0.037	0.037	0.038	0.039	0.040	0.040
4	AB	-0.022	-0.020	-0.004	-0.003	-0.002	-0.002	-0.006	-0.006	-0.001	-0.001	0.000	0.000
	AV	0.066	0.057	0.149	0.170	0.191	0.212	0.026	0.025	0.032	0.033	0.034	0.034
	AMSE	0.095	0.081	0.150	0.170	0.191	0.212	0.028	0.027	0.032	0.033	0.034	0.034
8	AB	-0.021	-0.015	-0.002	-0.002	-0.001	-0.001	-0.006	-0.005	0.000	0.000	0.000	0.000
	AV	0.064	0.044	0.184	0.225	0.271	0.328	0.024	0.022	0.032	0.033	0.034	0.035
	AMSE	0.088	0.058	0.184	0.225	0.271	0.328	0.025	0.023	0.032	0.033	0.034	0.035
Model	VARMA(1,1)	$T = 25$						$T = 100$					
$j =$		$\hat{\Psi}_j$	$\hat{\Gamma}_j$	$\hat{\Psi}_{j,j}^*$	$\hat{\Psi}_{j,j+1}^*$	$\hat{\Psi}_{j,j+2}^*$	$\hat{\Psi}_{j,j+3}^*$	$\hat{\Psi}_j$	$\hat{\Gamma}_j$	$\hat{\Psi}_{j,j}^*$	$\hat{\Psi}_{j,j+1}^*$	$\hat{\Psi}_{j,j+2}^*$	$\hat{\Psi}_{j,j+3}^*$
0	AB	-0.032	-0.032	-0.023	-0.020	-0.015	-0.011	-0.009	-0.009	-0.006	-0.005	-0.004	-0.003
	AV	0.170	0.170	0.184	0.180	0.187	0.192	0.046	0.046	0.047	0.047	0.047	0.048
	AMSE	0.178	0.178	0.190	0.182	0.189	0.193	0.047	0.047	0.047	0.047	0.047	0.048
1	AB	-0.035	-0.034	-0.021	-0.016	-0.012	-0.009	-0.009	-0.009	-0.005	-0.004	-0.003	-0.002
	AV	0.122	0.118	0.150	0.149	0.156	0.161	0.035	0.034	0.036	0.036	0.037	0.037
	AMSE	0.133	0.128	0.152	0.151	0.157	0.162	0.035	0.035	0.036	0.036	0.037	0.037
4	AB	-0.039	-0.035	-0.010	-0.008	-0.006	-0.005	-0.010	-0.010	-0.002	-0.001	-0.001	-0.001
	AV	0.104	0.088	0.181	0.178	0.188	0.196	0.030	0.029	0.034	0.034	0.034	0.034
	AMSE	0.116	0.098	0.181	0.178	0.188	0.197	0.031	0.029	0.034	0.034	0.034	0.034
8	AB	-0.035	-0.026	-0.004	-0.004	-0.003	-0.003	-0.010	-0.009	-0.001	0.000	0.000	0.000
	AV	0.103	0.071	0.280	0.265	0.302	0.367	0.029	0.027	0.036	0.036	0.036	0.037
	AMSE	0.113	0.076	0.280	0.265	0.302	0.367	0.029	0.027	0.036	0.036	0.036	0.037

Note: AB, AV and AMSE are defined in Eq. (4.7). The data generating process is defined in Eq. (4.2) with parameters provided in Eq. (4.5). The VMA(1) and the VAR(1) models have  $\Phi_1 = \mathbf{0}_{3 \times 3}$  and  $\Theta_1 = \mathbf{0}_{3 \times 3}$ , respectively. All results are based on a Monte Carlo size of 100,000.

Table 6: Average bias, variance and MSE of  $\hat{\Psi}_{j,h}^*$ ,  $\hat{\Psi}_j$  and  $\hat{\Gamma}_j$  for  $N = 5$

Model	VMA(1)	$T = 25$						$T = 100$					
$j =$		$\hat{\Psi}_j$	$\hat{\Gamma}_j$	$\hat{\Psi}_{j,j}^*$	$\hat{\Psi}_{j,j+1}^*$	$\hat{\Psi}_{j,j+2}^*$	$\hat{\Psi}_{j,j+3}^*$	$\hat{\Psi}_j$	$\hat{\Gamma}_j$	$\hat{\Psi}_{j,j}^*$	$\hat{\Psi}_{j,j+1}^*$	$\hat{\Psi}_{j,j+2}^*$	$\hat{\Psi}_{j,j+3}^*$
0	AB	-0.021	-0.021	-0.007	0.001	0.001	0.001	-0.006	-0.006	-0.002	0.000	0.000	0.000
	AV	0.114	0.114	0.124	0.132	0.134	0.135	0.030	0.030	0.031	0.031	0.031	0.031
	AMSE	0.116	0.116	0.124	0.132	0.134	0.135	0.030	0.030	0.031	0.031	0.031	0.031
1	AB	-0.025	-0.025	0.000	0.000	0.000	0.000	-0.006	-0.006	0.000	0.000	0.000	0.000
	AV	0.094	0.091	0.118	0.123	0.125	0.127	0.025	0.025	0.027	0.027	0.027	0.027
	AMSE	0.098	0.094	0.118	0.123	0.125	0.127	0.026	0.025	0.027	0.027	0.027	0.027
4	AB	-0.022	-0.019	0.000	0.000	0.000	0.000	-0.005	-0.005	0.000	0.000	0.000	0.000
	AV	0.088	0.074	0.148	0.158	0.164	0.168	0.024	0.023	0.027	0.027	0.027	0.027
	AMSE	0.091	0.076	0.148	0.158	0.164	0.168	0.024	0.023	0.027	0.027	0.027	0.027
8	AB	-0.022	-0.015	0.000	0.000	0.000	0.000	-0.006	-0.005	0.000	0.000	0.000	0.000
	AV	0.088	0.060	0.239	0.279	0.317	0.391	0.024	0.022	0.029	0.030	0.030	0.030
	AMSE	0.091	0.061	0.239	0.279	0.317	0.391	0.024	0.022	0.029	0.030	0.030	0.030
Model	VAR(1)	$T = 25$						$T = 100$					
$j =$		$\hat{\Psi}_j$	$\hat{\Gamma}_j$	$\hat{\Psi}_{j,j}^*$	$\hat{\Psi}_{j,j+1}^*$	$\hat{\Psi}_{j,j+2}^*$	$\hat{\Psi}_{j,j+3}^*$	$\hat{\Psi}_j$	$\hat{\Gamma}_j$	$\hat{\Psi}_{j,j}^*$	$\hat{\Psi}_{j,j+1}^*$	$\hat{\Psi}_{j,j+2}^*$	$\hat{\Psi}_{j,j+3}^*$
0	AB	-0.018	-0.018	-0.006	0.000	-0.002	-0.002	-0.005	-0.005	-0.001	0.000	-0.001	-0.001
	AV	0.081	0.081	0.087	0.090	0.089	0.090	0.021	0.021	0.022	0.022	0.022	0.022
	AMSE	0.082	0.082	0.088	0.091	0.089	0.090	0.021	0.021	0.022	0.022	0.022	0.022
1	AB	-0.020	-0.020	0.001	-0.002	-0.002	0.000	-0.005	-0.005	0.000	0.000	0.000	0.000
	AV	0.067	0.064	0.081	0.083	0.084	0.085	0.018	0.018	0.019	0.019	0.019	0.019
	AMSE	0.069	0.066	0.081	0.083	0.084	0.085	0.018	0.018	0.019	0.019	0.019	0.019
4	AB	-0.021	-0.019	-0.001	0.000	0.000	0.000	-0.005	-0.005	0.000	0.000	0.000	0.000
	AV	0.066	0.055	0.107	0.111	0.110	0.113	0.017	0.017	0.020	0.020	0.020	0.020
	AMSE	0.067	0.056	0.107	0.111	0.110	0.113	0.018	0.017	0.020	0.020	0.020	0.020
8	AB	-0.019	-0.013	0.000	0.000	0.000	-0.001	-0.005	-0.004	0.000	0.000	0.000	0.000
	AV	0.066	0.045	0.173	0.190	0.193	0.239	0.017	0.016	0.021	0.021	0.021	0.021
	AMSE	0.067	0.045	0.173	0.190	0.193	0.239	0.018	0.016	0.021	0.021	0.021	0.021
Model	VARMA(1,1)	$T = 25$						$T = 100$					
$j =$		$\hat{\Psi}_j$	$\hat{\Gamma}_j$	$\hat{\Psi}_{j,j}^*$	$\hat{\Psi}_{j,j+1}^*$	$\hat{\Psi}_{j,j+2}^*$	$\hat{\Psi}_{j,j+3}^*$	$\hat{\Psi}_j$	$\hat{\Gamma}_j$	$\hat{\Psi}_{j,j}^*$	$\hat{\Psi}_{j,j+1}^*$	$\hat{\Psi}_{j,j+2}^*$	$\hat{\Psi}_{j,j+3}^*$
0	AB	-0.047	-0.047	-0.028	-0.011	-0.010	-0.006	-0.012	-0.012	-0.007	-0.003	-0.002	-0.001
	AV	0.187	0.187	0.203	0.221	0.225	0.231	0.051	0.051	0.052	0.054	0.054	0.054
	AMSE	0.194	0.194	0.206	0.222	0.226	0.231	0.052	0.052	0.053	0.054	0.054	0.054
1	AB	-0.056	-0.055	-0.012	-0.010	-0.007	-0.002	-0.014	-0.014	-0.003	-0.002	-0.001	0.000
	AV	0.164	0.159	0.207	0.220	0.227	0.233	0.046	0.046	0.049	0.050	0.050	0.050
	AMSE	0.175	0.169	0.208	0.221	0.227	0.233	0.047	0.047	0.049	0.050	0.050	0.050
4	AB	-0.055	-0.048	-0.003	-0.001	-0.001	-0.001	-0.014	-0.014	0.000	0.000	0.000	0.000
	AV	0.148	0.124	0.257	0.280	0.290	0.303	0.042	0.040	0.048	0.049	0.049	0.049
	AMSE	0.157	0.131	0.257	0.280	0.290	0.303	0.043	0.041	0.048	0.049	0.049	0.049
8	AB	-0.048	-0.035	-0.001	0.000	-0.001	-0.001	-0.012	-0.011	0.000	0.000	0.000	0.000
	AV	0.148	0.101	0.400	0.478	0.527	0.640	0.042	0.038	0.052	0.053	0.053	0.053
	AMSE	0.155	0.104	0.400	0.478	0.527	0.640	0.042	0.039	0.052	0.053	0.053	0.053

Note: AB, AV and AMSE are defined in Eq. (4.7). The data generating process is defined in Eq. (4.2) with parameters provided in Eq. (4.6). The VMA(1) and the VAR(1) models have  $\Phi_1 = \mathbf{0}_{5 \times 5}$  and  $\Theta_1 = \mathbf{0}_{5 \times 5}$ , respectively. All results are based on a Monte Carlo size of 100,000.

Results in Tables 5 and 6 are largely consistent with our findings in the univariate case, with  $\hat{\Psi}_{j,h}^*$  having a smaller average bias which decreases as  $h$  increases at the cost of a larger variance and MSE in general. The average bias is very close to zero for the VMA(1) model with  $h > 1$  as expected, since  $\hat{\Psi}_{j,h}^*$  is exactly unbiased for the VMA( $h$ ) model. All estimators perform better as  $T$  increases, but the relative performances among  $\hat{\Psi}_{j,h}^*$ ,  $\hat{\Psi}_j$  and  $\hat{\Gamma}_j$  are qualitatively unchanged.

## 5 Further Discussions and Empirical Considerations

To our best knowledge, the usage of circular sample moment estimators on non-circular data is rare in the extant literature apart from the concurrent work of Li (2020). This could be due to its inferior finite sample performance relative to its conventional counterpart, as is demonstrated in our simulation study. Although we do not recommend to use the circular sample cross-covariance directly, we clearly show that its symmetric structure can be exploited to construct nearly unbiased cross-covariance estimators, which have a comparable performance to its conventional bias-corrected counterpart in a univariate setting but with a much simpler structure. As the cross-covariance matrix is in itself an important measure of association within a multivariate system, our proposed estimator provides a simple non-parametric estimator that is less distorted by the time series dependence structure within the system. For the readers who are interested in implementing our estimator or replicating our results, we provide MATLAB codes for the implementation of all bias-corrected estimators in the supporting information of the paper.

As an extension of this study, we expect that the  $\mathbf{A}$ -estimator, and hence the generalized  $\mathbf{A}$ -estimator, can also be formulated in the multivariate case using our framework. The multivariate generalized  $\mathbf{A}$ -estimator by construction nests our circular-based cross-covariance estimator, and is expected to have better finite sample properties based on our univariate simulation results. However, the corresponding mapping matrix in the multivariate case differs from  $\mathbf{A}$ , which requires individual inspection that is beyond the scope of this paper.

An important potential application of our estimator is to compute a bias-corrected long run variance-covariance matrix (e.g. Newey and West (1987); Andrews (1991)), which is used extensively in diagnostic tests of GMM estimation (see e.g. Hall (1996) and the reference therein). However, a major challenge here is to ensure the positive definiteness of the resulting estimates, which provides room for future research.

## 6 Conclusion

In this paper, we propose a simple nearly unbiased estimator for cross-covariances based on circular sample cross-covariance estimators. We show that this estimator is exactly unbiased under the VMA( $h$ )

model and is nearly unbiased when the VMA( $h$ ) structure provides a good approximation to the data generating process. Similar to the  $\mathbf{A}$ -estimator of [Vogelsang and Yang \(2016\)](#), our estimator is asymptotically equivalent to the conventional sample cross-covariance estimation, which allows a bias correction without sacrificing efficiency in large sample. We also propose a generalized  $\mathbf{A}$ -estimator that nests both our estimator and the  $\mathbf{A}$ -estimator in the univariate case, which is also asymptotically equivalent to the conventional sample autocovariance. Our simulation shows that in a univariate setting, our circular-based estimator performs almost as good as the  $\mathbf{A}$ -estimator in terms of bias, variance and MSE, while the generalized  $\mathbf{A}$ -estimator has overall the smallest MSE. Gains in MSE can be achieved when the autocovariance structure of the data decays slowly. For multivariate time series, we confirm our theoretical results that our estimator can correct the finite sample bias of cross-covariance estimators.

## Supporting Information

MATLAB codes for implementing the bias-corrected cross-covariance estimators and the generalized  $\mathbf{A}$ -estimator are provided, which can be found in the online version of this paper at the publisher's website.

## References

- Andrews, D. W. K. (1991). Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation. *Econometrica*, 59(3):817–858.
- Frutos, R. F. D. and Serrano, G. (2002). A generalized least squares estimation method for VARMA models. *Statistics*, 36(4):303–316.
- Hall, A. R. (1996). *Generalized Method of Moments: Advanced Text in Econometrics*, volume 100-101. Oxford University Press, New York.
- Henderson, H. V. and Searle, S. R. (1979). Vec and Vech Operators for Matrices, with Some Uses in Jacobians and Multivariate Statistics. *The Canadian Journal of Statistics/La Revue Canadienne de Statistique*, 7(1):65–81.
- Hosking, J. R. M. (1980). The Multivariate Portmanteau Statistic. *Journal of the American Statistical Association*, 75(371):602–608.
- Hosking, J. R. M. (1981). Equivalent Forms of the Multivariate Portmanteau Statistic. *Journal of the Royal Statistical Society: Series B (Methodological)*, 43(2):261–262.
- Kascha, C. (2012). A Comparison of Estimation Methods for Vector Autoregressive Moving-Average Models. *Econometric Reviews*, 31(3):297–324.



- Koreisha, S. and Pukkila, T. (1989). Fast Linear Estimation Methods for Vector Autoregressive Moving-Average Models. *Journal of Time Series Analysis*, 10(4):325–339.
- Koreisha, S. G. and Pukkila, T. M. (1988). Identification of Nonzero Elements in the Polynomial Matrices of Mixed Varma Processes. *Journal of the Royal Statistical Society: Series B (Methodological)*, 50(1):155.
- Li, W. K. and McLeod, A. I. (1981). Distribution of the Residual Autocorrelations in Multivariate ARMA Time Series Models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 43(2):231–239.
- Li, Y. (2020). Nearly unbiased estimation of sample skewness. *Economics Letters*, 192(109174).
- Mahdi, E. and Ian McLeod, A. (2012). Improved multivariate portmanteau test. *Journal of Time Series Analysis*, 33(2):211–222.
- Newey, W. K. and West, K. D. (1987). A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, 55(3):703.
- Okui, R. (2010). Asymptotically unbiased estimation of autocovariances and autocorrelations with long panel data. *Econometric Theory*, 26:1263–1304.
- Okui, R. (2011). Asymptotically unbiased estimation of autocovariances and autocorrelations for panel data with incidental trends. *Economics Letters*, 112(1):49–52.
- Politis, D. N. and Romano, J. P. (1992). A circular block-resampling procedure for stationary data. In *Exploring the limits of bootstrap. Wiley series in probability and mathematical analysis*. Wiley, New York.
- Priestley, M. B. (1981). *Spectral analysis and time series*. Academic Press, New York.
- Vogelsang, T. J. and Yang, J. (2016). Exactly/Nearly Unbiased Estimation of Autocovariances of a Univariate Time Series With Unknown Mean. *Journal of Time Series Analysis*, 37(6):723–740.
- Wei, W. W. S. (2019). *Multivariate Time Series Analysis and Applications*. Hoboken: John Wiley & Sons, Inc.
- Yang, J. and Vogelsang, T. J. (2018). Finite sample performance of a long run variance estimator based on exactly (almost) unbiased autocovariance estimators. *Economics Letters*, 165:21–27.

## A Technical Proofs

*Proof to Theorem 1.* Fix the notation  $\tilde{Y}_t = Y_t - \mathbb{E}[Y_t]$  and  $\check{Y} = \frac{1}{T} \sum_{t=1}^T \tilde{Y}_t$ . Starting with the first claim:

$$\begin{aligned} \mathbb{E}[\hat{\Psi}_0] &= \frac{1}{T} \sum_{t=1}^T \mathbb{E}[(\tilde{Y}_t - \check{Y})(\tilde{Y}_t - \check{Y})'] \\ &= \frac{1}{T} \sum_{t=1}^T \mathbb{E}[\tilde{Y}_t \tilde{Y}_t'] - \mathbb{E}[\check{Y} \check{Y}'] \\ &= \Gamma_0 - \mathbb{E}[\check{Y} \check{Y}']. \end{aligned} \tag{A.1}$$

We now examine the term  $\mathbb{E}[\check{Y} \check{Y}']$  carefully:

$$\check{Y} \check{Y}' = \frac{1}{T^2} \left( \sum_{t=1}^T \tilde{Y}_t \tilde{Y}_t' + \sum_{t=2}^T \sum_{j=1}^{t-1} \tilde{Y}_t \tilde{Y}_j' + \sum_{t=1}^{T-1} \sum_{j=t+1}^T \tilde{Y}_t \tilde{Y}_j' \right) \tag{A.2}$$

Taking expectation on both sides of the above equation yields:

$$\begin{aligned} \mathbb{E}[\check{Y} \check{Y}'] &= \frac{1}{T} \Gamma_0 + \frac{1}{T} \sum_{j=1}^{T-1} \frac{T-j}{T} (\Gamma_j + \Gamma_{-j}) \\ &= \frac{1}{T} \Gamma_0 + \frac{1}{T} K_T. \end{aligned} \tag{A.3}$$

Plugging the above into Eq. (A.1) yields claim 1 as desired. For claim 2:

$$\begin{aligned} \mathbb{E}[\hat{\Psi}_j] &= \frac{1}{T} \sum_{t=1}^T \mathbb{E}[(\tilde{Y}_t - \check{Y})(\tilde{Y}_{t-j} - \check{Y})'] \\ &= \frac{1}{T} \sum_{t=1}^T \mathbb{E}[\tilde{Y}_t \tilde{Y}_{t-j}'] - \mathbb{E}[\check{Y} \check{Y}'] \\ &= \frac{T-j}{T} \Gamma_j + \frac{j}{T} \Gamma_{T-j}' - \frac{1}{T} K_T - \frac{1}{T} \Gamma_0. \end{aligned} \tag{A.4}$$

This completes the proof.  $\square$

*Proof to Proposition 1.* For the first relation:

$$\hat{\Psi}_j = \frac{1}{T} \sum_{t=1}^T \hat{y}_t \hat{y}'_{t-j} = \frac{1}{T} \sum_{t=j+1}^T \hat{y}_t \hat{y}'_{t-j} + \frac{1}{T} \sum_{t=1}^j \hat{y}_t \hat{y}'_{T-j+t} = \hat{\Gamma}_j + \hat{\Gamma}'_{T-j}. \tag{A.5}$$

The second relation follows directly from the first one and that  $\hat{\Gamma}_j = \hat{\Gamma}'_{-j}$ . The third relation follows from:

$$\begin{aligned} \sum_{j=0}^{T-1} \hat{\Psi}_j &= \frac{1}{T} \sum_{j=0}^{T-1} \sum_{t=1}^T (Y_t - \bar{Y})(y_{Y-j} - \bar{Y})' \\ &= \frac{1}{T} \sum_{t=1}^T \sum_{j=0}^{T-1} (Y_t - \bar{Y})(Y_{t-j} - \bar{Y})' \\ &= \frac{1}{T} \sum_{t=1}^T (Y_t - \bar{Y}) \sum_{j=0}^{T-1} (Y_{t-j} - \bar{Y})' \\ &= \frac{1}{T} \sum_{t=1}^T (Y_t - \bar{Y}) \cdot \mathbf{0}'_{N \times 1} = \mathbf{0}_{N \times N}. \end{aligned} \tag{A.6}$$

This completes the proof. □

*Proof to Theorem 2.* We begin with the following linear system of equations which holds for any  $m$  and  $n$  under the VMA( $h$ ) assumption of  $Y$ :

$$\mathbb{E}[\widehat{\Psi}(m, n)] = \mathbf{H}_h \Gamma_h(m, n). \quad (\text{A.7})$$

This can be derived directly from Eq. (2.6) by setting  $\Gamma_j(m, n) = \Gamma_j(n, m) = 0$  for all  $j > h$ . Also, we have:

**Lemma 1.** For all  $h < \lfloor \frac{T}{2} \rfloor$ ,  $\mathbf{H}_h$  has full column rank such that  $\text{rank}(\mathbf{H}_h) = 2h + 1$ .

*Proof.* It suffices to prove the above lemma for the  $\tilde{h} = \lfloor \frac{T}{2} \rfloor - 1$  case, since for any  $h < \tilde{h}$ ,  $\mathbf{H}_h$  can be obtained by deleting  $2(\tilde{h} - h)$  columns from  $\mathbf{H}_{\tilde{h}}$ , which implies that  $\mathbf{H}_h$  must also have full column rank if  $\mathbf{H}_{\tilde{h}}$  is of full column rank. The dimension of  $\mathbf{H}_{\tilde{h}}$  depends on the parity of  $T$ , but the proof is identical. For conciseness, we prove the slightly more complicated case when  $T$  is odd:  $T = 2\tilde{h} + 3$ . In this case,  $\mathbf{H}_{\tilde{h}}$  is a  $(2\tilde{h} + 2)$ -by- $(2\tilde{h} + 1)$  matrix. Also, for the trivial case when  $T = 2, 3$ ,  $\mathbf{H}_{\tilde{h}}$  is a non-zero vector which is by construction of full rank.

We now prove the equivalent statement that  $\mathbf{H}'_{\tilde{h}}$  has full row rank by computing its reduced row echelon form (rref) explicitly. For a general matrix with a total of  $M$  rows, let us define the three elementary row operations as follows:

- Switching rows  $m$  and  $n$ :  $L_m \rightarrow L_n$ .
- Multiplying row  $m$  by a constant  $k \in \mathbb{R} \setminus \{0\}$ :  $kL_m \rightarrow L_m$ .
- Add a multiple of row  $n$  to row  $m$ :  $L_m + kL_n \rightarrow L_m$ , for some  $k \in \mathbb{R}$ .

For notational convenience, we also defined the following augmented row operation:

- Let  $\alpha(I) = \{\alpha(1), \alpha(2), \dots, \alpha(M)\}$  be a permutation of the row index  $I = \{1, 2, \dots, M\}$ . Permuting the rows from  $I$  to  $\alpha(I)$ :  $L_{\alpha(I)} \rightarrow L_I$ .

The above augmented operation is elementary in the sense that it can always be achieved by applying the switching row operation for a finite number of times. We now perform the following elementary row operations on  $\mathbf{H}'_{\tilde{h}}$ :

1. Define  $\alpha(I) = \{2, 3, \dots, \tilde{h} + 1, 1, 2\tilde{h} + 1, 2\tilde{h}, \dots, \tilde{h} + 2\}$ . Perform  $L_{\alpha(I)} \rightarrow L_I$ .

2. Perform the following row multiplication operations:

$$\begin{cases} \frac{T^2}{T-m}L_m \rightarrow L_m, & m \leq \tilde{h}, \\ TL_m \rightarrow L_m, & m = \tilde{h} + 1, \\ \frac{T^2}{T+\tilde{h}+1-m}L_m \rightarrow L_m, & m \geq \tilde{h} + 2. \end{cases} \quad (\text{A.8})$$

3. Perform  $L_m - L_{\tilde{h}+1} \rightarrow L_m, \forall m \neq \tilde{h} + 1$ .

4. Perform the following row multiplication operations:

$$\begin{cases} \frac{1}{T}L_m \rightarrow L_m, & m \neq \tilde{h} + 1, \\ -L_m \rightarrow L_m, & m = \tilde{h} + 1. \end{cases} \quad (\text{A.9})$$

5. Perform  $L_h - L_{\tilde{h}+1} \rightarrow L_{\tilde{h}+1}, \forall m \neq \tilde{h} + 1$ .

After the above straightforward operations, the resulting rref of  $\mathbf{H}'_{\tilde{h}}$  is the following  $(2\tilde{h}+1)$ -by- $(2\tilde{h}+2)$  matrix:

$$\text{rref}(\mathbf{H}'_{\tilde{h}}) = \begin{bmatrix} \mathbf{I}_{\tilde{h}+1} & \mathbf{0}_{\tilde{h} \times (\tilde{h}+1)} \\ \mathbf{0}_{\tilde{h} \times (\tilde{h}+1)} & \mathbf{I}_{\tilde{h}+1} \end{bmatrix}. \quad (\text{A.10})$$

Clearly the above matrix has full row rank, which implies the lemma for the odd  $T$  case. We note that for the case when  $T$  is even,  $\text{rref}(\mathbf{H}'_{\tilde{h}}) = \mathbf{I}_{2\tilde{h}+1}$  by applying the same set of operations. As the above set of operations can always be performed for any  $T \geq 4$ , the proof is complete.  $\square$

Lemma 1 implies that  $\mathbf{H}'_h \mathbf{W} \mathbf{H}_h$  must be positive-definite and hence invertible for any  $h < \lfloor \frac{T}{2} \rfloor$  and a positive-definite  $\mathbf{W}$ . The following result is then straightforward:

$$\mathbb{E}[(\mathbf{H}'_h \mathbf{W} \mathbf{H}_h)^{-1} \mathbf{H}'_h \mathbf{W} \hat{\Psi}(m, n)] = (\mathbf{H}'_h \mathbf{W} \mathbf{H}_h)^{-1} \mathbf{H}'_h \mathbf{W} \mathbf{H}_h \Gamma_h(m, n) = \Gamma_h(m, n). \quad (\text{A.11})$$

Note that for the above result to hold, it is crucial that  $\mathbf{W}$  is non-random. This proves the unbiasedness of  $\hat{\Psi}_h^*(m, n)$ . We note that the result also holds for the case  $m = n$ , however it may deliver different estimates for  $\Gamma_j(m, n)$  and  $\Gamma_j(n, m)$ . The consistency of  $\hat{\Psi}_h^*(m, n)$  follows directly from the consistency of  $\hat{\Psi}(m, n)$  and the continuous mapping theorem, and the proof is complete.  $\square$

*Proof to Corollary 1.* When  $h = 0$  and  $\mathbf{W} = \mathbf{I}_{T-1}$ ,  $\mathbf{H}_0 = -\frac{1}{T}\boldsymbol{\nu}_{(T-1) \times 1}$ , where  $\boldsymbol{\nu}_{m \times n}$  is an  $m$ -by- $n$  matrix of ones. We therefore see that:

$$(\mathbf{H}'_0 \mathbf{H}_0)^{-1} \mathbf{H}'_0 \hat{\Psi}(m, n) = \frac{T^2}{T-1} \cdot -\frac{1}{T}\boldsymbol{\nu}_{(T-1) \times 1} \hat{\Psi}(m, n) = \frac{T}{T-1} \hat{\Psi}_0(m, n), \quad (\text{A.12})$$

where we have used the relationship  $-\boldsymbol{\nu}_{(T-1) \times 1} \hat{\Psi}(m, n) = \hat{\Psi}_0(m, n)$  according to Proposition 1. This completes the proof.  $\square$

*Proof to Proposition 2.* When  $Y$  follows a  $\text{VMA}(\infty)$  model, we always have:

$$\mathbb{E}[\widehat{\Psi}(m, n)] = \mathbf{H}_h \boldsymbol{\Gamma}_h(m, n) + \boldsymbol{\epsilon}(m, n), \quad (\text{A.13})$$

which can be derived equation-by-equation from Eq. (2.6). It is then clear that the proposition can be proved by multiplying both sides of the above equation by  $(\mathbf{H}'_h \mathbf{W} \mathbf{H}_h)^{-1} \mathbf{H}'_h \mathbf{W}$ . This completes the proof.  $\square$

*Proof to Proposition 3.* We prove  $\mathbf{H}'_h = \mathbf{H}'_h \mathbf{H}_h \mathbf{R}_h$ , which implies  $(\mathbf{H}'_h \mathbf{H}_h)^{-1} \mathbf{H}'_h = \mathbf{R}_h$  and hence the proposition. Note that  $\mathbf{H}'_h \mathbf{H}_h$  is always invertible due to Lemma 1. Starting with the fact that  $\mathbf{H}_h \mathbf{R}_h$  has a block-diagonal structure:

$$\mathbf{H}_h \mathbf{R}_h = \begin{bmatrix} \mathbf{I}_h & \mathbf{0}_{h \times (T-2h-1)} & \mathbf{0}_{h \times h} \\ \mathbf{0}'_{h \times (T-2h-1)} & \boldsymbol{\nu}_{(T-2h-1) \times 1} \otimes \mathbf{v}'_h & \mathbf{0}'_{h \times (T-2h-1)} \\ \mathbf{0}_{h \times h} & \mathbf{0}_{h \times (T-2h-1)} & \mathbf{I}_h \end{bmatrix}, \quad (\text{A.14})$$

where  $\otimes$  is the tensor product. This is derived by direct calculation of the matrix multiplication. We now write  $\mathbf{H}'_h = (\mathbf{H}_h^{(1)}, \mathbf{H}_h^{(2)}, \mathbf{H}_h^{(3)})$ , where  $\mathbf{H}_h^{(1)}$  and  $\mathbf{H}_h^{(3)}$  cover the first and last  $h$  columns of  $\mathbf{H}'_h$ . It is therefore clear that  $\mathbf{H}'_h = \mathbf{H}'_h \mathbf{H}_h \mathbf{R}_h$  is equivalent to:

$$\mathbf{H}_h^{(2)} = \mathbf{H}_h^{(2)} \boldsymbol{\nu}_{(T-2h-1) \times 1} \otimes \mathbf{v}'_h. \quad (\text{A.15})$$

Observe that  $\mathbf{H}_h^{(2)}$  has the following structure:

$$\mathbf{H}_h^{(2)} = \begin{bmatrix} -\frac{1}{T} \boldsymbol{\nu}'_{(T-2h-1) \times 1} \\ -\frac{T-1}{T^2} \boldsymbol{\nu}'_{(T-2h-1) \times 1} \\ \vdots \\ -\frac{T-h}{T^2} \boldsymbol{\nu}'_{(T-2h-1) \times 1} \\ -\frac{T-1}{T^2} \boldsymbol{\nu}'_{(T-2h-1) \times 1} \\ \vdots \\ -\frac{T-h}{T^2} \boldsymbol{\nu}'_{(T-2h-1) \times 1} \end{bmatrix}. \quad (\text{A.16})$$

We also have the following identity:

$$\boldsymbol{\nu}'_{(T-2h-1) \times 1} \boldsymbol{\nu}_{(T-2h-1) \times 1} \otimes \mathbf{v}'_h = \boldsymbol{\nu}'_{(T-2h-1) \times 1}. \quad (\text{A.17})$$

Equations (A.16) and (A.17) imply Eq. (A.15). This completes the proof.  $\square$

*Proof to Corollary 2.* It suffices to notice that the columns of  $\widehat{\Psi}^\circ$  and  $\widehat{\Gamma}_h^\circ$  are simply  $\widehat{\Psi}(m, n)$  and  $\boldsymbol{\Gamma}_h(m, n)$ , and the corollary follows by principles of matrix multiplication. This completes the proof.  $\square$

*Proof to Corollary 3.* For the case  $j \geq 1$ , let us write  $\mathbf{R}_h(j)$  as the  $(j+1)$ -th row of  $\mathbf{R}_h$ . Using Proposition 3, we see that:

$$\text{vec}(\hat{\Psi}_j^*)' = \mathbf{R}_h(j)\hat{\Psi}^\circ = \frac{T}{T-j}\text{vec}(\hat{\Psi}_j)' - \frac{T}{(T-j)(T-2h-1)} \sum_{i=h+1}^{T-h-1} \text{vec}(\hat{\Psi}_i)'. \quad (\text{A.18})$$

For the case with  $j = 0$ , we note that:

$$\text{vec}(\hat{\Psi}_0^*)' = \mathbf{R}_h(0)\hat{\Psi}^\circ = - \sum_{i=1}^h (\text{vec}(\hat{\Psi}_i) + \text{vec}(\hat{\Psi}_{T-i})) - \frac{T-2h}{T-2h-1} \sum_{i=h+1}^{T-h-1} \text{vec}(\hat{\Psi}_i)'. \quad (\text{A.19})$$

Proposition 1 implies that:

$$- \sum_{i=1}^{T-1} \text{vec}(\hat{\Psi}_i) = \text{vec}(\hat{\Psi}_0). \quad (\text{A.20})$$

Plugging the above into Eq. (A.19) yields:

$$\text{vec}(\hat{\Psi}_0^*)' = \text{vec}(\hat{\Psi}_0)' - \frac{1}{T-2h-1} \sum_{i=h+1}^{T-h-1} \text{vec}(\hat{\Psi}_i)'. \quad (\text{A.21})$$

Reversing the vectorization, we see that this is the identical to Eq. (A.19) with  $j = 0$ . Finally, again using Eq. (A.20):

$$- \sum_{i=h+1}^{T-h-1} \text{vec}(\hat{\Psi}_i)' = \sum_{i=-h}^h \text{vec}(\hat{\Psi}_i)'. \quad (\text{A.22})$$

Substituting the above into Eq. (A.19) and reversing the vectorization yield the desired result. The unbiasedness and consistency follow directly from Proposition 2. This completes the proof.  $\square$

*Proof to Proposition 4.* From Corollary 3, we have:

$$\hat{\Psi}_{0,h}^* = \hat{\Psi}_0 + \frac{1}{T-2h-1} \sum_{i=-h}^h \hat{\Psi}_i. \quad (\text{A.23})$$

Since  $\hat{\Psi}_j = \hat{\Gamma}_j + \hat{\Gamma}_{T-j}$ , we can rewrite  $\hat{\Psi}_{0,h}^*$  in the following form:

$$\hat{\Psi}_{0,h}^* = \sum_{j=-(T-1)}^{T-1} k(j)\hat{\Gamma}_j, \quad (\text{A.24})$$

where:

$$k(j) = \begin{cases} \frac{T-2h}{T-2h-1}, & j = 0, \\ \frac{1}{T-2h-1}, & |j| \in [1, T-1] \setminus [h+1, T-h-1], \\ 0, & |j| \in [h+1, T-h-1]. \end{cases} \quad (\text{A.25})$$

As a standard result of spectral density estimation (see e.g. Priestley (1981)),  $\hat{\Psi}_{0,h}^*$  is positive definite with probability 1 if the following condition is met:

$$K(x) = \frac{1}{2\pi} \sum_{j=-(T-1)}^{T-1} k(j)e^{ijx} \geq 0, \quad \forall x \in [-\pi, \pi], \quad (\text{A.26})$$

where  $i = \sqrt{-1}$  is the imaginary unit. Note that  $K(x)$  satisfies the following identity:

$$K(x) = \frac{1}{2\pi} \left( 1 + \frac{\sum_{j=-(T-1)}^{T-1} e^{ijx} - \sum_{j=-(T-h-1)}^{T-h-1} e^{ijx} + \sum_{j=-h}^h e^{ijx}}{T - 2h - 1} \right). \quad (\text{A.27})$$

Note that the Dirichlet kernel is defined via the following well-known identity:

$$D_h(x) = \frac{1}{2\pi} \sum_{j=-h}^h e^{ijx} = \frac{\sin((2h+1)x/2)}{2\pi \sin(x/2)}. \quad (\text{A.28})$$

Plugging the above into Eq. (A.27), we arrive at:

$$K(x) = \frac{1}{2\pi} \left( 1 + \frac{2\pi(D_{T-1}(x) - D_{T-h-1}(x) + D_h(x))}{T - 2h - 1} \right) \geq 0, \quad \forall x \in [-\pi, \pi], \quad (\text{A.29})$$

which implies Eq. (3.13), and the proof is complete.  $\square$

*Proof to Theorem 3.* From Proposition 1 it is clear that  $\hat{\Psi}_j$  and  $\hat{\Gamma}_j$  are asymptotically equivalent for any finite  $j$ . To prove the theorem, we firstly write:

$$\hat{\Psi}_h(m, n) = \{ \{ \hat{\Psi}_j(m, n) \}'_{j=0:h}, \{ \hat{\Psi}_j(n, m) \}'_{j=1:h} \}', \quad \hat{\Gamma}_h(m, n) = \{ \{ \hat{\Gamma}_j(m, n) \}'_{j=0:h}, \{ \hat{\Gamma}_j(n, m) \}'_{j=1:h} \}'. \quad (\text{A.30})$$

We propose to show that  $\hat{\Psi}_h^*(m, n)$ , which are the columns of  $\hat{\Psi}_h^*$ , is asymptotically equivalent to  $\hat{\Psi}_h(m, n)$  for any  $m$  and  $n$ . Since  $\hat{\Psi}_h(m, n)$  is asymptotically equivalent to  $\hat{\Gamma}_h(m, n)$  for any fixed  $h$ , which are the columns of  $\hat{\Gamma}_h^\circ$ , the theorem follows as each column of  $\hat{\Psi}_h^*$  is asymptotically equivalent to the corresponding column of  $\hat{\Gamma}_h^\circ$ . Expanding  $\hat{\Psi}_h^*(m, n)$  explicitly yields:

$$\hat{\Psi}_h^*(m, n) = \mathbf{R}_h \hat{\Psi}(m, n) = \hat{\Psi}_h(m, n) + \begin{bmatrix} -\mathbf{v}'_h \hat{\Psi}_{h+}(m, n) \\ \frac{\hat{\Psi}_1(m, n)}{T-1} - \frac{T}{T-1} \mathbf{v}'_h \hat{\Psi}_{h+}(m, n) \\ \vdots \\ \frac{h\hat{\Psi}_h(m, n)}{T-h} - \frac{T}{T-h} \mathbf{v}'_h \hat{\Psi}_{h+}(m, n) \\ \frac{\hat{\Psi}_1(n, m)}{T-1} - \frac{T}{T-1} \mathbf{v}'_h \hat{\Psi}_{h+}(m, n) \\ \vdots \\ \frac{h\hat{\Psi}_h(n, m)}{T-h} - \frac{T}{T-h} \mathbf{v}'_h \hat{\Psi}_{h+}(m, n), \end{bmatrix}. \quad (\text{A.31})$$

where  $\hat{\Psi}_{h+}(m, n) = \{ \hat{\Psi}_j \}'_{j=h+1:T-h-1}$ . By the absolute summability condition in Assumption 1, it must hold that  $\mathbf{v}'_{(T-2h-1) \times 1} \hat{\Psi}_{h+}(m, n) = \mathcal{O}_p(1)$  for every  $m$  and  $n$  in the limit. Therefore we see that, for any finite  $h$ ,  $\frac{h}{T-h} = \mathcal{O}(T^{-1})$ , which implies  $\sqrt{T}(\hat{\Psi}_{h+}(m, n) - \hat{\Psi}_h(m, n)) = \mathcal{O}_p(T^{-1}) - \mathcal{O}_p(T^{-0.5}) = o_p(1)$ , and hence the required asymptotic equivalence. This completes the proof.  $\square$

*Proof of Corollary 4.* Using the result in Corollary 3, we can easily construct the  $(h+1)$ -by- $(h+1)$  matrix  $\tilde{\mathbf{M}}_h$  that satisfies  $\hat{\psi}_h^* = \tilde{\mathbf{M}}_h \hat{\psi}_h$  by collecting coefficients from Eq. (3.10), where  $\hat{\psi}_h = \{ \hat{\psi}_j \}_{j=0:h}$ . Denote  $\tilde{\mathbf{M}}_h = \{ M_{i,j} \}_{i=1:h+1, j=1:h+1}$ , this gives:

$$\tilde{M}_{i,j} = \begin{cases} \frac{T}{T-i+1} (\mathbb{1}_{\{i=1\}} + \frac{1}{T-2h-1}), & i \in [1, h+1], j = 1 \\ \frac{T}{T-i+1} (\mathbb{1}_{\{i=j\}} + \frac{2}{T-2h-1}), & i \in [1, h+1], j \in [2, h+1]. \end{cases} \quad (\text{A.32})$$

Now from Proposition 1, we have  $\hat{\psi}_j = \hat{\gamma}_j + \hat{\gamma}_{T-j}$ , which implies that the  $(h+1+j)$ -th column of  $\mathbf{M}_h$  is just the  $(T-j+1)$ -th column of  $\widetilde{\mathbf{M}}_h$  for  $j \in [1, h]$ , and the proof is complete.  $\square$

*Proof of Proposition 5.* The proposition is a direct result of the following identity derived in Equation (2) of Vogelsang and Yang (2016):

$$\mathbb{E}[\widehat{\boldsymbol{\gamma}}] = \mathbf{A}\boldsymbol{\gamma}. \quad (\text{A.33})$$

Therefore, by removing the last  $T-h-1$  elements from  $\boldsymbol{\gamma}$  and for any increasing sequence  $J_q \subseteq \{0, \dots, T-1\}$ , the following relation holds due to the principle of matrix multiplication:

$$\mathbb{E}[\widehat{\boldsymbol{\gamma}}_{J_q}] = \mathbf{A}_{J_q}\boldsymbol{\gamma}_h. \quad (\text{A.34})$$

Under the condition that  $\mathbf{A}_{J_q}$  has full column rank,  $\mathbf{A}'_{J_q}\mathbf{W}\mathbf{A}_{J_q}$  is guaranteed to be invertible, and the proof therefore follows analogously as the proof of Theorem 2. This completes the proof.  $\square$

*Proof of Theorem 4.* Choose some fixed  $h \leq T-2$ . We start with the following decomposition of  $\mathbf{A}_{J_T}$ , which is a direct consequence of Lemma 1 in Vogelsang and Yang (2016):

$$\mathbf{A}_{J_T} = \mathbf{C}_{h+1} - T^{-1}\mathbf{B}_T, \quad (\text{A.35})$$

where  $\mathbf{C}_{h+1} = \{\mathbf{I}_{h+1}, \mathbf{0}_{(h+1) \times (T-h-1)}\}'$  satisfies  $\mathbf{C}'_{h+1}\mathbf{C}_{h+1} = \mathbf{I}_{h+1}$ ,  $\mathbf{B}_T = \mathbf{B} + \mathcal{O}(T^{-1})$  and  $\mathbf{B}$  is some constant  $T$ -by- $(h+1)$  matrix that does not depend on  $T$ . We therefore have:

$$\begin{aligned} \mathbf{A}'_{J_T}\mathbf{A}_{J_T} &= \mathbf{I}_{h+1} - 2T^{-1}\mathbf{C}'_{h+1}\mathbf{B}_T + T^{-2}\mathbf{B}'_T\mathbf{B}_T \\ &= \mathbf{I}_{h+1} + T^{-1}\mathbf{U}_T\mathbf{B}_T, \end{aligned} \quad (\text{A.36})$$

where  $\mathbf{U}_T = -2\mathbf{C}'_{h+1} + T^{-1}\mathbf{B}'_T$ . The Woodbury matrix identity suggests that:

$$\begin{aligned} (\mathbf{A}'_{J_T}\mathbf{A}_{J_T})^{-1}\mathbf{A}'_{J_T} &= (\mathbf{I}_{h+1} - T^{-1}\mathbf{U}_T(\mathbf{A}'_{J_T}\mathbf{A}_{J_T})^{-1}\mathbf{B}_T)\mathbf{A}'_{J_T} \\ &= \mathbf{A}'_{J_T} - T^{-1}\mathbf{U}_T(\mathbf{A}'_{J_T}\mathbf{A}_{J_T})^{-1}\mathbf{B}_T\mathbf{A}'_{J_T} \\ &= \mathbf{C}'_{h+1} + \mathcal{O}(T^{-1}). \end{aligned} \quad (\text{A.37})$$

Using the identity  $\mathbf{C}'_{h+1}\widehat{\boldsymbol{\gamma}} = \widehat{\boldsymbol{\gamma}}_h$ , we see that:

$$\sqrt{T}(\widehat{\boldsymbol{\gamma}}_{h, J_T}^* - \widehat{\boldsymbol{\gamma}}_h) = \sqrt{T}((\mathbf{A}'_{J_T}\mathbf{A}_{J_T})^{-1}\mathbf{A}'_{J_T}\widehat{\boldsymbol{\gamma}} - \widehat{\boldsymbol{\gamma}}_h) = \mathcal{O}(T^{0.5}) \cdot \mathcal{O}(T^{-1}) \cdot \mathcal{O}_p(1) = o_p(1), \quad (\text{A.38})$$

as desired, and the proof is complete.  $\square$