



# Asymmetric vector moving average models: estimation and testing

Jan G. De Gooijer<sup>1</sup> 

Received: 18 February 2020 / Accepted: 2 December 2020 / Published online: 1 February 2021  
© The Author(s) 2021

## Abstract

We propose the class of asymmetric vector moving average (asVMA) models. The asymmetry of these models is characterized by different MA filters applied to the components of vectors of lagged positive and negative innovations. This allows for a detailed investigation of the interrelationships among past model innovations of different sign. We derive some covariance matrix properties of an asVMA model under the assumption of Gaussianity. Related to this, we investigate the global invertibility condition of the proposed model. The paper also introduces a maximum likelihood estimation procedure and a multivariate Wald-type test statistic for symmetry versus the alternative of asymmetry. The finite-sample performance of the proposed multivariate test is studied by simulation. Furthermore, we devise an exploratory test statistic based on lagged sample cross-bicovariance estimates. The estimation and testing procedures are used to uncover asymmetric effects in two US growth rates, and in three US industrial prices.

**Keywords** Asymmetries · Cross-bicovariance estimates · Multivariate · Test performance · Wald-type test statistic

## 1 Introduction

It is widely believed that there is an asymmetric inertia in many major economic time series, often attributed to differences in time series dynamics in periods of business-cycle contraction and expansion. This asymmetric behavior has been documented by, for instance, Wecker (1981) for US industrial prices. Brännäs and De Gooijer (1994) and Elwood (1998) find empirical evidence of asymmetry in US real GNP growth rates. Some more recent works include Gonzalo and Martínez (2006) for US GNP

---

✉ Jan G. De Gooijer  
j.g.degooijer@uva.nl

<sup>1</sup> Amsterdam School of Economics, University of Amsterdam, PO Box 15867, 1001 Amsterdam, NJ, The Netherlands

data, and Taştan (2017) for Turkish real GNP and industrial production index series. Further related papers are by Brännäs and de Luna (1998), Brännäs and Ohlsson (1999), Nebeling and Salish (2017), and Sáfadi and Morettin (1998).

A common feature of the above studies is the use of univariate asymmetric moving average (asMA) models and their variants. The dynamics of these models respond to innovations with one of two different rules according to whether the innovation is positive or negative, and hence induces asymmetry. Obviously, the univariate structure of an asMA model offers limited information about the asymmetry in a data generating process (DGP) because dynamic interrelationship between several variables is ignored. This calls for a more flexible multivariate (vector) dynamic specification with similar features as the univariate asMA model.

The purpose of the paper is to introduce and study the class of asymmetric vector moving average (asVMA) models. The asVMA model may be viewed as an extension of the univariate asMA model proposed by Wecker (1981). Roughly speaking, the asymmetric effect in an asVMA model is characterized by different MA filters applied to the components of vectors of lagged positive and negative innovations. This allows for a detailed investigation of the interrelationships among model innovations of different sign. From an empirical standpoint, there are sound reasons for using vector nonlinear models to detect asymmetries in macroeconomic time series; see, e.g., Atanasova (2003), Keating (2000), and Weisse (1999).

The rest of the paper is organized as follows. In Sect. 2, we introduce the asVMA model and discuss some covariance properties. Related to this, we briefly investigate the global invertibility of the proposed model. Section 3 is on estimation and testing. First, we describe the log-likelihood function. Next, we propose a multivariate Wald-type test statistic for testing an asVMA against a linear vector MA model. We evaluate the finite-sample performance of the proposed Wald-type test statistic in a Monte Carlo simulation study. In Sect. 4, we devise an exploratory test statistic based on lagged sample cross-bicovariance estimates. Section 5 contains two illustrative applications. A summary is given in Sect. 6. All proofs are relegated to an “Appendix”.

## 2 Asymmetric vector moving average model and properties

### 2.1 Model

Consider an  $m$ -dimensional stochastic process  $\mathbf{Y}_t = (Y_{1t}, \dots, Y_{mt})'$ . Let  $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \dots, \varepsilon_{mt})'$  be an  $m$ -dimensional i.i.d. white noise process with  $m \times 1$  mean vector  $\mathbf{0}_m$  and  $m \times m$  positive definite matrix  $\boldsymbol{\Sigma}_\varepsilon$ , independent of  $\mathbf{Y}_t$ . Assume that the dynamic relationships in  $\{\mathbf{Y}_t, t \in \mathbb{Z}\}$  are represented through a linear vector filter of positive innovations  $\boldsymbol{\varepsilon}_t^+ = (\varepsilon_{1t}^+, \dots, \varepsilon_{mt}^+)'$  and a linear filter of negative innovations  $\boldsymbol{\varepsilon}_t^- = (\varepsilon_{1t}^-, \dots, \varepsilon_{mt}^-)'$  which satisfy, respectively,

$$\begin{aligned}\boldsymbol{\varepsilon}_t^+ &= (\max\{0, \varepsilon_{1t}\}, \dots, \max\{0, \varepsilon_{mt}\})' = (I(\varepsilon_{1t} \geq 0)\varepsilon_{1t}, \dots, I(\varepsilon_{mt} \geq 0)\varepsilon_{mt})', \\ \boldsymbol{\varepsilon}_t^- &= (\min\{0, \varepsilon_{1t}\}, \dots, \min\{0, \varepsilon_{mt}\})' = (I(\varepsilon_{1t} < 0)\varepsilon_{1t}, \dots, I(\varepsilon_{mt} < 0)\varepsilon_{mt})',\end{aligned}$$

where  $I(\cdot)$  denotes the indicator function. Then an  $m$ -dimensional asymmetric vector moving average process of order  $(q, q)$ , denoted by asVMA( $q, q$ ), is defined as

$$\begin{aligned}
 \mathbf{Y}_t &= \boldsymbol{\varepsilon}_t + \sum_{v=1}^q \mathbf{B}_v^+ \boldsymbol{\varepsilon}_{t-v}^+ + \sum_{v=1}^q \mathbf{B}_v^- \boldsymbol{\varepsilon}_{t-v}^- \\
 &= \boldsymbol{\varepsilon}_t + \sum_{v=1}^q \mathbf{B}_v^+ \boldsymbol{\varepsilon}_{t-v} + \sum_{v=1}^q (\mathbf{B}_v^- - \mathbf{B}_v^+) \boldsymbol{\varepsilon}_{t-v}^-, \tag{1}
 \end{aligned}$$

where  $\mathbf{B}_v^\pm = [b_{r,s,v}^\pm]_{r,s=1}^m$  ( $v = 1, \dots, q$ ) are  $m \times m$  parameter matrices.

We see that in (1), each component of the vector process  $\boldsymbol{\varepsilon}_t$  is transformed by separate MA filters and the asymmetry in  $\{\mathbf{Y}_t, t \in \mathbb{Z}\}$  depends on the sign of the innovations. Moreover, in (1) we assume that each component of  $\boldsymbol{\varepsilon}_t$  has the option of moving into one of two possible directions, above and below a fixed value zero. Note that by introducing a suitable number of zero parameter matrices, the order  $q$  can be different for  $\boldsymbol{\varepsilon}_t^+$  and  $\boldsymbol{\varepsilon}_t^-$ . When  $(\mathbf{B}_v^- - \mathbf{B}_v^+) = \mathbf{0}_{m \times m}$ , (1) reduces to an  $m$ -dimensional VMA( $q$ ) process in ‘‘standard form’’. Also note that for  $m = 1$ , model (1) corresponds to the univariate asMA( $q$ ) process proposed by Wecker (1981), i.e.,  $Y_t = \varepsilon_t + \sum_{v=1}^q b_v^+ I(\varepsilon_{t-v} \geq 0) \varepsilon_{t-v} + \sum_{v=1}^q b_v^- I(\varepsilon_{t-v} < 0) \varepsilon_{t-v}$ . For one-step ahead forecasts of stock returns, Koutmos (1999) used a variant of this model with the conditional mean given by an asMA(1) model and the conditional standard deviation given by a threshold GARCH(1, 1) model. Guay and Scaillet (2003) modified the univariate asMA( $q$ ) to allow for contemporaneous asymmetric effects.

### 2.2 Invertibility

Forecasting with an asVMA model is, in principle, the same as producing forecasts from a VMA model provided the model is invertible. For each univariate component of  $\mathbf{Y}_t$ , the empirical invertibility of (1) can be checked by Monte Carlo simulation; see, e.g., De Gooijer and Brännäs (1995). Their approach can be easily extended to the multivariate case. Alternatively, we can use the following global invertibility condition (Niglio and Vitale 2013)

$$|\lambda(\boldsymbol{\Psi}^+)|^p |\lambda(\boldsymbol{\Psi}^-)|^{1-p} < 1, \tag{2}$$

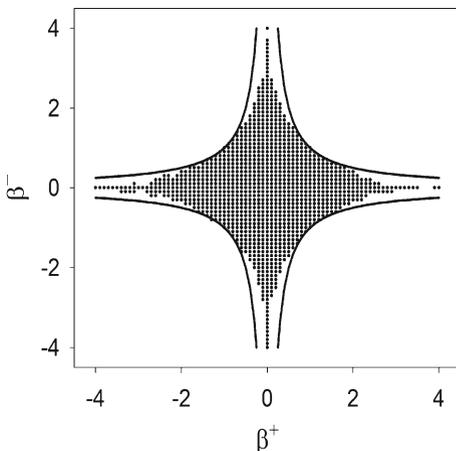
where  $\boldsymbol{\Psi}^\pm$  is an  $mq \times mq$  matrix defined by

$$\boldsymbol{\Psi}^\pm = \begin{pmatrix} \mathbf{B}_1^\pm & & & \mathbf{B}_q^\pm \\ & \dots & & \\ & & \mathbf{0}_{(m-1)q \times mq} & \end{pmatrix}.$$

Here,  $\lambda(\cdot)$  is the maximum eigenvalue of a square matrix and  $p = \mathbb{E}[I(\mathbf{1}'_m \boldsymbol{\varepsilon}_{t-1}^+ \geq 0)]$  with  $\mathbf{1}_m$  an  $m \times 1$  vector consisting of ones.

As an illustration, consider a two-dimensional asVMA(1,1) model with  $\mathbf{B}_1^+ = -\mathbf{B}_1^-$  and with the values of  $b_{12,1}^\pm$  fixed at zero. The reason for choosing this particular model will become clear in Sect. 2.3. Using (2) with  $p = 1/2$ , we noticed that the shape of the

**Fig. 1** Scatter plot of the global invertibility region for the pair of parameters  $(\beta^+, \beta^-) \equiv (b_{rr,1}^+, b_{ss,1}^-)$  ( $r, s = 1, 2$ ) of a two-dimensional asVMA(1, 1) model with  $\mathbf{B}_1^+ = -\mathbf{B}_2^-$  and with  $b_{12,\pm}^\pm = 0$ . The black solid lines represent the condition  $|\beta^+||\beta^-| < 1$



global invertibility regions are the same for all pairs of parameters  $(b_{rr,1}^+, b_{ss,1}^-)$  ( $r, s = 1, 2$ ). Figure 1 shows a graphical representation where the elements of  $\mathbf{B}_1^\pm$  are taken in the range  $[-4, 4]$  using a step-size of 0.1. It is interesting to see that the region indicates invertibility over a wide range of parameter values. In fact, a simple approximation to the global invertibility region is given by the condition  $|b_{rr,1}^+||b_{ss,1}^-| < 1$ . This result follows from approximating the values at the boundaries of the invertibility–non-invertibility region by nonlinear functions of the parameter combinations.

**2.3 Covariance properties**

For a linear stationary vector time series process  $\{\mathbf{Y}_t, t \in \mathbb{Z}\}$ , the mean vector, variance and cross-covariance matrices provide useful summary information on the strength and direction between its components. Given a particular VMA model, these theoretical properties are well known. By contrast, explicit expressions for the mean and covariance of a general asVMA process are more difficult to derive.

Consider as an illustrative example the following  $m$ -dimensional asVMA(1, 1) process

$$\mathbf{Y}_t = \boldsymbol{\varepsilon}_t + \mathbf{B}^+ \boldsymbol{\varepsilon}_{t-1} + (\mathbf{B}^- - \mathbf{B}^+) (I(\varepsilon_{1,t-1} < 0) \varepsilon_{1,t-1}, \dots, I(\varepsilon_{m,t-1} < 0) \varepsilon_{m,t-1})', \tag{3}$$

where we suppressed the subscript  $v$  in  $\mathbf{B}_v^\pm$ . The next proposition summarizes some properties of (3).

**Proposition** *Let  $\{\mathbf{Y}_t, t \in \mathbb{Z}\}$  follow the process in (3) with  $\{\boldsymbol{\varepsilon}_t\} \stackrel{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}_m, \boldsymbol{\Sigma}_\varepsilon)$ . Then the unconditional mean vector, the unconditional variance-covariance matrix, and the unconditional lag- $\ell$  cross-covariance matrix are given by, respectively,*

$$\mathbb{E}(\mathbf{Y}_t) = (2\pi \mathbf{1}'_m \boldsymbol{\Sigma}_\varepsilon \mathbf{1}_m)^{-1/2} (\mathbf{B}^- - \mathbf{B}^+) \boldsymbol{\Sigma}_\varepsilon \mathbf{1}_m,$$

$$\begin{aligned} \text{Var}(\mathbf{Y}_t) &= \boldsymbol{\Sigma}_\varepsilon + \frac{1}{2} \left( \mathbf{B}^- \boldsymbol{\Sigma}_\varepsilon (\mathbf{B}^-)' + \mathbf{B}^+ \boldsymbol{\Sigma}_\varepsilon (\mathbf{B}^+)' \right) - \mathbb{E}(\mathbf{Y}_t) \mathbb{E}(\mathbf{Y}_t)', \\ \boldsymbol{\Gamma}(\ell) &= [\gamma_{ij}(\ell)]_{i,j=1}^m = \text{Cov}(\mathbf{Y}_t, \mathbf{Y}_{t-\ell}) = \begin{cases} \frac{1}{2} (\mathbf{B}^- + \mathbf{B}^+) \boldsymbol{\Sigma}_\varepsilon, & \ell = 1, \\ \mathbf{0}_{m \times m}, & \ell > 1. \end{cases} \end{aligned}$$

We see that unlike the  $m$ -dimensional VMA(1) process where  $\mathbb{E}(\mathbf{Y}_t) = \mathbf{0}_m$ , the mean vector of the asVMA(1, 1) process in (3) is a function of the parameter matrices  $\mathbf{B}^-$  and  $\mathbf{B}^+$  and is in general a nonzero vector. However, the  $m$ -dimensional asVMA(1, 1) process like the  $m$ -dimensional VMA(1) process, has the property that  $\boldsymbol{\Gamma}(\ell) = \mathbf{0}_{m \times m}$  for  $\ell > 1$ . Thus, it will be hard to distinguish between both processes on the basis of the sample estimate of  $\boldsymbol{\Gamma}(\ell)$ . Also, setting  $\mathbf{B}^- = -\mathbf{B}^+$ , it follows from the Proposition that  $\boldsymbol{\Gamma}(\ell) = \mathbf{0}_{m \times m}$  for all lags  $\ell \geq 1$ . When this asymmetry in parameter configuration occurs it is impossible to distinguish an  $m$ -dimensional asVMA(1, 1) process from an  $m$ -dimensional Gaussian white noise process using the lag- $\ell$  cross-covariance matrix.

Following the proofs in the ‘‘Appendix’’, the results in the Proposition can be extended to asVMA processes with  $q > 1$ . Then the same confusing situation occurs as above, i.e., it is impossible to distinguish between an asVMA process with  $\mathbf{B}_v^- = -\mathbf{B}_v^+$  and an  $m$ -dimensional Gaussian white noise process, based on the unconditional lag- $\ell$  cross-covariance matrix alone. In Sect. 3.2, we overcome this problem by introducing the proposed Wald-type test statistic.

### 3 Estimation and testing

For ease of notation, and without loss of generality, we assume throughout this section that the asVMA model order is fixed at  $q$ . In Sect. 5 we present estimation results with some lags excluded from the fitted model specification.

#### 3.1 Likelihood

Assume that  $\{\mathbf{Y}_t\}_{t=1}^T$  is generated by the  $m$ -dimensional asVMA( $q, q$ ) model in (1) with a constant vector term  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_m)'$  included on the right-hand side. For the conditional likelihood approach, we also assume that the initial observations  $\mathbf{Y}_0, \mathbf{Y}_{-1}, \dots, \mathbf{Y}_{1-T}$  are fixed, and  $\boldsymbol{\varepsilon}_0 = \boldsymbol{\varepsilon}_{-1} = \dots = \boldsymbol{\varepsilon}_{1-T} = \mathbf{0}_m$ . This assumption does not affect the asymptotic properties of the test statistic developed in Sect. 3.2. We define the  $T \times m$  matrices  $\mathbf{Y} = ((\mathbf{Y}_1 - \boldsymbol{\mu}), \dots, (\mathbf{Y}_T - \boldsymbol{\mu}))'$ , and  $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_T)'$ . In addition, using the vectorizing operation ‘‘vec’’ which forms a vector from a matrix by stacking the columns of the matrix one underneath the other, we define the  $mT \times 1$  vectors

$$\mathbf{y} = \text{vec}(\mathbf{Y}'), \quad \mathbf{e} = \text{vec}(\boldsymbol{\varepsilon}')$$

as well as the  $m^2 \times 1$  parameter vector  $\boldsymbol{\beta}_v^{(i)} = \text{vec}(\mathbf{B}_v^{(i)})$  ( $i = 1, 2; v = 1, \dots, q$ ), introducing the notation  $\mathbf{B}_v^{(1)} \equiv \mathbf{B}_v^-$  and  $\mathbf{B}_v^{(2)} \equiv \mathbf{B}_v^+$ . Next, we introduce the  $T \times T$

lag matrix  $\mathbf{L}$  with ones on the (sub)diagonal directly below the principal diagonal and zeros elsewhere, with  $(\mathbf{L}^v \otimes \mathbf{I}_m)\mathbf{e} = (\mathbf{0}'_m, \dots, \mathbf{0}'_m, \boldsymbol{\varepsilon}'_1, \dots, \boldsymbol{\varepsilon}'_{T-v})'$  ( $v = 1, \dots, q$ ). Let  $\mathbf{D}_v = (\mathbf{B}_v^{(1)} - \mathbf{B}_v^{(2)})$ . Then the model in (1) can be written as

$$\begin{aligned} \mathbf{y} &= \mathbf{e} + \sum_{v=1}^q (\mathbf{L}^v \otimes \mathbf{B}_v^{(2)})\mathbf{e} + \sum_{v=1}^q (\mathbf{Q}\mathbf{L}^v \otimes \mathbf{D}_v)\mathbf{e} \\ &= (\boldsymbol{\Theta}^{(1)} + \boldsymbol{\Theta}^{(2)})\mathbf{e}, \end{aligned} \tag{4}$$

where  $\boldsymbol{\Theta}^{(1)} = (\mathbf{I}_T \otimes \mathbf{I}_m) + \sum_{v=1}^q (\mathbf{L}^v \otimes \mathbf{B}_v^{(2)})$  and  $\boldsymbol{\Theta}^{(2)} = \sum_{v=1}^q (\mathbf{Q}\mathbf{L}^v \otimes \mathbf{D}_v)$  are two  $mT \times mT$  matrices, and  $\mathbf{Q} = \text{diag}\{[I(\varepsilon_{11} < 0), \dots, I(\varepsilon_{m1} < 0)], \dots, [I(\varepsilon_{1T} < 0), \dots, I(\varepsilon_{mT} < 0)]\}$  is a  $T \times T$  diagonal matrix.

Define the  $(2qm^2 + m) \times 1$  parameter vector

$$\boldsymbol{\theta} = (\boldsymbol{\mu}', (\boldsymbol{\beta}_1^{(1)})', \dots, (\boldsymbol{\beta}_q^{(1)})', (\boldsymbol{\beta}_1^{(2)})', \dots, (\boldsymbol{\beta}_q^{(2)})')'. \tag{5}$$

Then on the assumption of normality of the  $\boldsymbol{\varepsilon}_t$  and since  $\mathbf{e} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(\mathbf{0}_{mT}, \mathbf{I}_T \otimes \boldsymbol{\Sigma}_\varepsilon)$ , the log-likelihood function, apart from an additive constant term, is given by

$$\begin{aligned} \ell(\boldsymbol{\theta}) &= \sum_{t=1}^T \ell_t(\boldsymbol{\theta}) = -\frac{T}{2} \log |\boldsymbol{\Sigma}_\varepsilon| - \frac{1}{2} \sum_{t=1}^T \boldsymbol{\varepsilon}'_t \boldsymbol{\Sigma}_\varepsilon^{-1} \boldsymbol{\varepsilon}_t \\ &= -\frac{T}{2} \log |\boldsymbol{\Sigma}_\varepsilon| - \frac{1}{2} \mathbf{e}' (\mathbf{I}_T \otimes \boldsymbol{\Sigma}_\varepsilon^{-1}) \mathbf{e}, \end{aligned} \tag{6}$$

where  $\mathbf{e} = \boldsymbol{\Theta}^{-1}\mathbf{y}$  with  $\boldsymbol{\Theta} = \boldsymbol{\Theta}^{(1)} + \boldsymbol{\Theta}^{(2)}$ , and assuming the inverse exist. For a fixed parameter vector  $\boldsymbol{\theta}$ , it is clear that maximization of (6) with respect to  $\boldsymbol{\Sigma}_\varepsilon$  yields  $\widehat{\boldsymbol{\Sigma}}_\varepsilon = \sum_{t=1}^T \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t / T$ .

For a vector autoregressive moving average model with Gaussian distributed errors, Reinsel et al. (1992, Sect. 2) derived an expression for the vector of partial derivatives (also called gradient or score vector) of the log-likelihood function with respect to the parameter vector. Their result carries over to an asVMA( $q, q$ ) process. In particular the expression for the partial derivatives of the log-likelihood function  $\ell(\boldsymbol{\theta})$  is given by

$$\begin{aligned} \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &= -\frac{\partial \mathbf{e}'}{\partial \boldsymbol{\theta}} (\mathbf{I}_T \otimes \boldsymbol{\Sigma}_\varepsilon^{-1}) \mathbf{e} \\ &= \mathbf{Z}' (\boldsymbol{\Theta}')^{-1} (\mathbf{I}_T \otimes \boldsymbol{\Sigma}_\varepsilon^{-1}) \mathbf{e}, \end{aligned} \tag{7}$$

where  $\mathbf{Z} = [(\mathbf{L}\boldsymbol{\varepsilon} \otimes \mathbf{I}_m), \dots, (\mathbf{L}^q \boldsymbol{\varepsilon} \otimes \mathbf{I}_m)]$  is an  $mT \times (2qm^2 + m)$  matrix. In practice, Eq. (7) needs to be solved by an iterative numerical optimization procedure. Associated to this, it is often useful to have a convenient expression for the  $2qm^2 \times 2qm^2$  Hessian matrix  $\mathcal{H}(\boldsymbol{\theta})$  of second partial derivatives of  $\ell(\boldsymbol{\theta})$ . When  $T \rightarrow \infty$ , it is well known that  $\mathcal{H}(\boldsymbol{\theta})$  can be approximated by the outer product of the gradient vector, which is

equivalent to

$$\mathcal{H}(\boldsymbol{\theta}) \approx \left(\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right) \left(\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)' = \mathbf{Z}'(\boldsymbol{\Theta}')^{-1}(\mathbf{I}_T \otimes \boldsymbol{\Sigma}_\varepsilon^{-1})\boldsymbol{\Theta}^{-1}\mathbf{Z}. \tag{8}$$

In Sect. 3.2, we use the inverse of  $\mathcal{H}(\boldsymbol{\theta})$  as an approximation of the variance-covariance matrix of the vector of parameter estimates.

### 3.2 Wald-type test statistic

Let  $\boldsymbol{\beta} = ((\boldsymbol{\beta}^{(1)})', (\boldsymbol{\beta}^{(2)})')'$  denote the  $2qm^2 \times 1$  vector of parameters of an  $m$ -dimensional asVMA( $q, q$ ) process, excluding the constant term  $\boldsymbol{\mu}$ . It is apparent from the notations introduced in Sect. 3.1 that the problem of testing the null hypothesis of symmetry is equivalent to testing the restriction  $\boldsymbol{\beta}^{(1)} = \boldsymbol{\beta}^{(2)}$ . A convenient test statistic can be obtained as follows. Let  $\mathbf{R}$  denote a known restriction matrix of dimension  $qm^2 \times (2qm^2 + m)$  such that  $\mathbf{R}\boldsymbol{\theta} = \mathbf{r}$  with  $\mathbf{r}$  a  $qm^2$ -vector of restricted parameters. Next, from the partition  $\mathbf{R} = (\mathbf{R}_1 : \mathbf{R}_2)$ , where  $\mathbf{R}_1$  is a  $qm^2 \times m$  matrix of zeros and  $\mathbf{R}_2$  is a  $qm^2 \times 2qm^2$  matrix, the problem becomes one of testing the null hypothesis

$$\mathbb{H}_0 : \mathbf{R}_2\boldsymbol{\beta} = \mathbf{0}_{qm^2} \text{ versus } \mathbb{H}_1 : \mathbf{R}_2\boldsymbol{\beta} \neq \mathbf{0}_{qm^2}. \tag{9}$$

Let  $\widehat{\boldsymbol{\theta}}$  be the vector of parameter estimates of  $\boldsymbol{\theta}$  under  $\mathbb{H}_0$ , and  $\mathcal{H}^{-1}(\widehat{\boldsymbol{\theta}})$  the estimate of the corresponding covariance matrix evaluated under the null. Then the Wald-type (W) test statistic is given by

$$W_T = \widehat{\boldsymbol{\beta}}' \mathbf{R}'_2 [\mathbf{R} \mathcal{H}^{-1}(\widehat{\boldsymbol{\theta}}) \mathbf{R}']^{-1} \mathbf{R}_2 \widehat{\boldsymbol{\beta}}, \tag{10}$$

with  $\widehat{\boldsymbol{\beta}}$  denoting the unrestricted estimator of  $\boldsymbol{\beta}$ . Under  $\mathbb{H}_0$ , and as  $T \rightarrow \infty$ ,  $W_T \xrightarrow{D} \chi^2_{qm^2}$ . Any consistent estimate of  $\mathcal{H}(\boldsymbol{\theta})$  will lead to a different variant of the Wald-type test statistic.

**Remark 1** The asymptotic distribution of the restricted estimator  $\widehat{\boldsymbol{\theta}}$  can be established using a central limit theorem for martingale difference sequences (see, e.g., White 2001, Ch. 5), and is given by

$$T^{1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{D} \mathcal{N}_{2qm^2+m}(\mathbf{0}, (\mathbf{R}'\mathbf{V}\mathbf{R})^{-1}), \tag{11}$$

where  $\mathbf{V} = \lim_{T \rightarrow \infty} T^{-1} \mathbb{E}(-\partial^2 \ell(\boldsymbol{\theta}) / \partial \mathbf{r} \partial \mathbf{r}')$ . Given (11), the asymptotic distribution of the Wald-type test statistic (10) follows from White (2001, Thm. 4.31(ii)).

**Remark 2** Let  $\{q_{rs}^\pm\}_{r,s=1}^m$  denote the set of lag orders corresponding to the positive (negative) innovations of an  $m$ -dimensional asVMA model. Then, testing for symmetry implies  $q_{rs}^- = q_{rs}^+ \equiv q_{rs}$ . In this case the  $W_T$  is asymptotically  $\chi^2$  distributed with  $\sum_{r=1}^m \sum_{s=1}^m q_{rs}$  degrees of freedom.

**Table 1** Size of the Wald-type test statistic  $W_T$  at nominal size 0.05 for DGP-1 ( $m = 2$ ). Number of simulations is 1000

$\delta$	$T = 250$	$T = 500$	$T = 1000$	$T = 2000$
-0.4	0.053	0.057	0.048	0.052
-0.2	0.054	0.042	0.041	0.053
0	0.052	0.044	0.052	0.058
0.2	0.050	0.042	0.047	0.059
0.4	0.063	0.056	0.049	0.052

### 3.3 Test performance of $W_T$

To study the empirical size of  $W_T$ , we consider a two-dimensional asVMA(1, 1) (DGP-1) model with  $\mathbf{B}^\pm = \begin{pmatrix} \delta & -\delta \\ -\delta & \delta \end{pmatrix}$ , where  $\delta = 0, \pm 0.2, \pm 0.4$ , and  $\{\boldsymbol{\varepsilon}_t\} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(\mathbf{0}_2, \mathbf{I}_2)$ . To avoid the effect of any starting-up transient on the generated process, a prior part consisting of 500 observations was discarded from each series. Based on 1000 Monte Carlo replications, Table 1 shows the observed sizes of the Wald-type test statistic at nominal size 0.05 for  $T = 250, 500, 1000$  and 2000. The table reveals that  $W_T$  is generally well sized. It is also apparent that the observed sizes are about the same for positive and negative values of  $\delta$ .

In addition, we also studied the size of the Wald-type test statistic for a three-dimensional asVMA(1,1) model with  $\{\boldsymbol{\varepsilon}_t\} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(\mathbf{0}_3, \mathbf{I}_3)$  with the parameter matrix given by DGP-4 below. Under the  $\mathbb{H}_0: \mathbf{B}^+ = \mathbf{B}^-$  the size (5% nominal level) of the test statistic is 0.091 ( $T = 500$ ), 0.051 ( $T = 1000$ ), and 0.46 ( $T = 2000$ ) for  $\delta = 0$ . For  $|\delta| \geq 0.2$ , however, we noticed serious size distortions. The size improves slightly for  $T = 2000$ . Nevertheless, caution is needed when interpreting the results of the Wald-type test when the sample size is relatively small.

To study the empirical power of  $W_T$ , we employ three  $m$ -dimensional asVMA(1, 1) ( $m = 2, 3$ ) DGPs with the following parameter configurations,

$$\text{DGP-2: } \mathbf{B}^- = \begin{pmatrix} 0.5 & 0.2 \\ 0.4 & 0.3 \end{pmatrix}, \mathbf{B}^+ = \begin{pmatrix} 0.7 & -0.3 \\ -0.2 & -0.5 \end{pmatrix},$$

$$\text{DGP-3: } \mathbf{B}^- = \begin{pmatrix} 0.7 & 0 \\ 0.3 & 0.5 \end{pmatrix}, \mathbf{B}^+ = -\mathbf{B}^-,$$

$$\text{DGP-4: } \mathbf{B}^- = \begin{pmatrix} \delta & -\delta & 0 \\ -\delta & \delta & -\delta \\ 0 & -\delta & \delta \end{pmatrix}, \mathbf{B}^+ = -\mathbf{B}^-, \quad (\delta = 0.2, 0.4, 0.6, 0.8),$$

where in all cases  $\{\boldsymbol{\varepsilon}_t\} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(\mathbf{0}_m, \mathbf{I}_m)$ . Note, all DGPs do not have a constant vector term  $\boldsymbol{\mu}$ .

Table 2 shows the empirical rejection frequencies of the  $W_T$  test statistic at a 5% nominal significance level. The power of the test statistic is reasonable to good, irrespective of the dimension  $m$  and the sample size  $T$ . This also applies to the special case  $\mathbf{B}^+ = -\mathbf{B}^-$  (DGP-3 and DGP-4) which we discussed in Sect. 2.3.

**Table 2** Empirical probability of rejecting symmetry by the  $W_T$  test statistic when the DGP is given by the three  $m$ -dimensional asVMA(1, 1) models ( $m = 2, 3$ ) specified in Sect. 3.3

DGP	$m$	$\delta$	$T = 200$	$T = 500$	$T = 1000$
2	2		0.796	0.863	0.879
			0.732	0.802	0.857
		0.2	0.581	0.661	0.745
4	3	0.4	0.684	0.712	0.751
		0.6	0.784	0.795	0.802
		0.8	0.794	0.796	0.857

Nominal size is 5% and number of simulations is 1000

**Table 3** Parameter estimates and their corresponding standard deviations (in parentheses) for DGP-2 and DGP-3

$b_{rs}^\pm$	DGP-2 ( $m = 2$ )			$b_{rs}^\pm$	DGP-3 ( $m = 2$ )		
	$T = 200$	$T = 500$	$T = 1000$		$T = 200$	$T = 500$	$T = 1000$
0.5	0.503 (0.121)	0.500 (0.071)	0.500 (0.048)	0.7	0.705 (0.118)	0.704 (0.067)	0.701 (0.043)
0.2	0.215 (0.120)	0.205 (0.070)	0.201 (0.048)	0	0.003 (0.090)	0.003 (0.055)	0.003 (0.037)
0.4	0.412 (0.137)	0.406 (0.076)	0.403 (0.054)	0.3	0.306 (0.113)	0.302 (0.069)	0.301 (0.047)
0.3	0.306 (0.142)	0.304 (0.081)	0.301 (0.057)	0.5	0.503 (0.123)	0.496 (0.074)	0.501 (0.053)
0.7	0.715 (0.122)	0.707 (0.068)	0.702 (0.048)	-0.7	-0.717 (0.090)	-0.711 (0.052)	-0.699 (0.034)
-0.2	-0.208 (0.136)	-0.202 (0.079)	-0.201 (0.053)	0	-0.006 (0.090)	-0.005 (0.055)	-0.001 (0.038)
-0.3	-0.304 (0.114)	-0.304 (0.070)	-0.299 (0.049)	-0.3	-0.308 (0.115)	-0.306 (0.067)	-0.299 (0.047)
-0.5	-0.519 (0.147)	-0.508 (0.084)	-0.505 (0.059)	-0.5	-0.512 (0.106)	-0.503 (0.065)	-0.503 (0.042)

Number of simulations is 1000

Table 3 shows parameter estimates and their corresponding standard deviations for DGP-2 and DGP-3. In all cases the performance of the estimation method is good, i.e., both the bias and standard deviation of  $\hat{\theta}$  decreases as  $T$  increases. This empirical result is in agreement with the asymptotic result in (11).

## 4 Two exploratory test statistics

### 4.1 asMA model

To investigate the possible asymmetry of an asMA model ( $m = 1$ ), Welsh and Jernigan (1983) introduced the lag  $\ell$  bivariate function, defined as  $\tilde{\gamma}(\ell) = \text{Cov}(Y_{t-\ell}^2, Y_t)$ . They proposed testing the null hypothesis  $\mathbb{H}_0: \tilde{\gamma}(\ell) = 0$ . For  $\mathbb{E}(Y_t) = 0$ , an unbiased and consistent estimator of  $\tilde{\gamma}(\ell)$  is given by  $\hat{\gamma}(\ell) = (T - \ell)^{-1} \sum_{t=\ell+1}^T Y_{t-\ell}^2 Y_t$ . The

authors showed that under the null hypothesis the standardized bicovariance

$$Q(\ell) = \sum_{t=\ell+1}^T Y_{t-\ell}^2 Y_t / \sqrt{3(T-\ell)} \tag{12}$$

has an asymptotic  $\mathcal{N}(0, 1)$  distribution. For an asMA(1) process,  $Q(\ell)$  has good power as demonstrated by Brännäs et al. (1998).

**4.2 asVMA model**

A natural extension of  $\tilde{\gamma}(\ell)$  is the lag  $\ell$  cross-bicovariance function, defined as  $\tilde{\gamma}_{ij}(\ell) = \mathbb{E}[Y_{i,t-\ell}^2 Y_{j,t}]$  ( $i, j = 1, \dots, m; \ell \in \mathbb{N}^+$ ). However, unlike the correlation matrix used for the identification of linear vector time series processes, the resulting cross-bicovariance matrix is not symmetric with respect to the principal diagonal since  $\tilde{\gamma}_{ij}(\ell) \neq \tilde{\gamma}_{ji}(\ell)$ . To obtain a symmetric matrix, we first define the following matrices

$$\mathbf{\Gamma}^{(1)}(\ell) = [\gamma_{uv}^{(1)}]_{u,v=1}^m = \begin{cases} \tilde{\gamma}_{uv}(\ell), & u \leq v, \\ \tilde{\gamma}_{vu}(\ell), & u > v, \end{cases} \quad \mathbf{\Gamma}^{(2)}(\ell) = [\gamma_{uv}^{(2)}]_{u,v=1}^m = \begin{cases} \tilde{\gamma}_{vu}(\ell), & u \leq v, \\ \tilde{\gamma}_{uv}(\ell), & u > v. \end{cases} \tag{13}$$

Then, a symmetrized lag- $\ell$  cross-bicovariance matrix is given by

$$\tilde{\mathbf{\Gamma}}(\ell) = \frac{1}{2}(\mathbf{\Gamma}^{(1)}(\ell) + \mathbf{\Gamma}^{(2)}(\ell)). \tag{14}$$

Note that the principal diagonal of  $\tilde{\mathbf{\Gamma}}(\ell)$  is composed of the elements  $\tilde{\gamma}_{uu}(\ell)$  ( $u = 1, \dots, m$ ). The off-diagonals above and below the principal diagonal of  $\tilde{\mathbf{\Gamma}}(\ell)$  are composed of the elements  $(\tilde{\gamma}_{uv}(\ell) + \tilde{\gamma}_{vu}(\ell))/2$  ( $u \leq v; u, v = 1, \dots, m$ ), respectively.

Using (14), the null- and alternative hypotheses of interest are

$$\mathbb{H}_0: \tilde{\mathbf{\Gamma}}(\ell) = [\tilde{\gamma}_{ij}(\ell)]_{i,j=1}^m = \mathbf{0}_{m \times m} \text{ versus } \mathbb{H}_1: \tilde{\mathbf{\Gamma}}(\ell) \neq \mathbf{0}_{m \times m}, \quad (\ell = 1, \dots, q), \tag{15}$$

where  $q \in \mathbb{Z}^+$  is a prescribed constant integer. Let  $\hat{\mathbf{\Gamma}}_m(\ell) = [Z_{ij}(\ell)]_{i,j=1}^m$  denote an estimator of  $\tilde{\mathbf{\Gamma}}(\ell)$ , where  $Z_{ij}(\ell) = \sum_{t=\ell+1}^T Y_{i,t-\ell}^2 Y_{j,t} / \sqrt{3(T-\ell)}$ . Then, an exploratory test statistic for testing  $\mathbb{H}_0$  can be based on the squared Frobenius norm. That is, the test statistics is given by

$$Q_m(\ell) = \text{Tr}(\hat{\mathbf{\Gamma}}_m'(\ell)\hat{\mathbf{\Gamma}}_m(\ell)), \\ = \left( \sum_{i=1}^m Z_{ii}^2(\ell) + \frac{1}{2} \sum_{i \neq j}^m (Z_{ij}(\ell) + Z_{ji}(\ell))^2 \right), \quad (\ell = 1, \dots, q). \tag{16}$$

Large values of  $Q_m(\ell)$  indicate that  $\mathbb{H}_0$  should be rejected.

**Table 4** Empirical power of  $Q_m(1)$  when the vector time process follows two  $m$ -dimensional asVMA(1, 1) models ( $m = 2, 3$ ) for sample sizes  $T = 50, 100, 200$

$\delta$	DGP-6 ( $m = 2$ )			DGP-6 ( $m = 3$ )		
	$T = 50$	$T = 100$	$T = 200$	$T = 50$	$T = 100$	$T = 200$
0.1	0.072	0.088	0.172	0.134	0.268	0.593
0.2	0.184	0.347	0.719	0.534	0.865	0.998
0.3	0.371	0.704	0.973	0.873	0.997	1.000
0.4	0.583	0.898	0.999	0.985	1.000	1.000
0.5	0.743	0.970	1.000	0.999	1.000	1.000

Nominal size is 5%, number of bootstrap replicates  $B = 500$ , and number of simulations 1000

Given (16), it is easily verified that under  $\mathbb{H}_0$  the asymptotic distribution of  $Q_m(\ell)$  is characterized by the sum of two uncorrelated random variables, i.e.,  $Q_m(\ell) = (X_1 + X_2)$ . Here, the distribution of  $X_1$  is  $\chi^2(1)$  multiplied by a constant  $(m + 1)/2$ . Then, the random variable  $((m + 1)/2)\chi^2(1)$  has a gamma distribution  $\Gamma(k, \theta)$  with shape parameter  $k = 1/2$  and scale parameter  $\theta = (m + 1)$ . The distribution of  $X_2$  is  $(m - 1)/2$  times the product of two uncorrelated  $\mathcal{N}(0, 1)$  random variables, say  $Z$ , where the exact probability density function of  $Z$  is given by the well-known result  $f(z) = (1/\pi)K_0(z)$  with  $K_0(\cdot)$  the modified Bessel function of the second kind of order zero. Clearly, for  $m = 1$  the null distribution of  $Q_1(\ell)$  is  $\chi^2(1)$ . For  $m \geq 2$ , however, the null distribution of  $Q_m(\ell)$  is untractable. In that case, we use a stationary bootstrap scheme, with automatic block-length selection; see Politis and White (2004) and Patton et al. (2007).

### 4.3 Test performance of $Q_m(\ell)$

To study the empirical power of  $Q_m(\ell)$ , we employ two  $m$ -dimensional asVMA(1, 1) ( $m = 2, 3$ ) DGPs with the following parameter configurations

$$\begin{aligned} \text{DGP-5: } \mathbf{B}^- = -\mathbf{B}^+ &= \begin{pmatrix} \delta & -\delta \\ -\delta & \delta \end{pmatrix}, \quad \mathbf{\Sigma}_1 = \begin{pmatrix} 1 & 0.2 \\ 0.2 & 1 \end{pmatrix}, \quad \mathbf{\Sigma}_2 = \begin{pmatrix} 1 & 0.3 \\ 0.3 & 1 \end{pmatrix}, \\ \text{DGP-6: } \mathbf{B}^- = -\mathbf{B}^+ &= \begin{pmatrix} \delta & -\delta & 0 \\ -\delta & \delta & -\delta \\ 0 & -\delta & \delta \end{pmatrix}, \quad \mathbf{\Sigma}_1 = \begin{pmatrix} 1 & 0.2 & 0 \\ 0.2 & 1 & 0.2 \\ 0 & 0.2 & 1 \end{pmatrix}, \quad \mathbf{\Sigma}_2 = \begin{pmatrix} 1 & 0.3 & 0 \\ 0.3 & 1 & 0.3 \\ 0 & 0.3 & 1 \end{pmatrix}, \end{aligned}$$

where  $\delta = 0.1, \dots, 0.5$ , and  $\{\varepsilon_t\} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(\mathbf{0}_m, \mathbf{I}_m)$ .

For a 5% nominal significance level, Table 4 shows the empirical power of  $Q_m(\ell)$  at time-lag  $\ell = 1$  with respect to three different sample sizes  $T$  and five parameter values  $\delta$ . It is seen that  $Q_m(1)$  has good power for all dimensions  $m$ . For fixed values of  $\delta$ , the power improves with increasing sample size  $T$ . Note that for  $\delta = 0.1$  both DGPs are close to a vector Gaussian white noise process. But still the rejection rates are satisfactory. As expected, in all cases the empirical size (not shown here) is close to the nominal size.

### 5 Illustrative applications

No general model selection strategy has been developed to determine the most appropriate asVMA model in practice. We found it convenient to tackle this problem in the following way. First, based on the values of the test statistic  $Q(\ell)$  the best-fitted univariate asMA model is obtained for each time series separately. Next, we used these specifications as initial polynomial estimates for the diagonal elements of the associated asVMA model with low order polynomial specifications for the off-diagonal elements of the vector model. This provides an initial guess of the structure of the asVMA model, which can be fine-tuned using a suitable model order selection criterion (e.g.,  $AIC(q) = \ln |\widehat{\Sigma}_\varepsilon| + 2qm^2/T$ ) and by diagnostic checking the residuals. In addition, the pattern of the lag  $\ell$  cross-bicovariance matrix  $\widehat{\Gamma}_m(\ell)$  and the corresponding test statistic  $Q_m(\ell)$  may suggest directions of model improvement. However, in the interest of parsimony, we recommend that the initial asVMA model should be kept simple. Also, deleting asVMA model parameters which are small compared with their standard errors, may provide a better understanding of the DGP under study.

#### 5.1 GDP and CPI

As a first illustration, we use quarterly US real GDP (seasonally adjusted; not inflation adjusted) and quarterly US consumer price index (CPI) total all items (seasonally adjusted), covering the period 1960(i)–2017(iv). In particular, we employ the GDP growth rate, i.e.,  $Y_{1t} = \ln(GDP_t/GDP_{t-1})$  and the inflation rate  $Y_{2t} = \ln(CPI_t/CPI_{t-1})$  ( $t = 2, \dots, 332$ ). Our analysis starts with presenting results for the exploratory test statistics discussed in Sect. 4.

Table 5 reports values of the test statistics  $Q(\ell)$  and  $Q_2(\ell)$  ( $\ell = 1, \dots, 10$ ). For series  $Y_{1t}$ ,  $Q(\ell)$  has no values exceeding the 95% confidence limits at lags 1–10. For series  $Y_{2t}$ , however, significant values of  $Q(\ell)$  are at lags 1–7 and 9–10. On the other hand, the univariate version of the Wald-type test statistic rejects the null hypothesis of symmetry for both series. In Table 5 the numbers within parentheses are the  $p$ -values of the test statistic  $Q_2(\ell)$ . We see that at lags 1–6, the null hypothesis in (15) is rejected at the 5% significance level. This strongly suggest that an asVMA model should be entertained with parameter matrices specified at the first 6 lags, and perhaps also at lag 10.

In our further analysis, we employed AIC to search over different asVMA specifications. The “best” fitted model is a two-dimensional asVMA given by

$$\begin{aligned} \widehat{\mu} &= \begin{pmatrix} 0.007 \\ (4.597) \\ 0.011, \\ (8.240) \end{pmatrix}, \widehat{\mathbf{B}}_1^- = \begin{pmatrix} -0.151 & 0.426 \\ (-0.981) & (1.996) \\ 0.063 & 0.639 \\ (0.454) & (4.351) \end{pmatrix}, \widehat{\mathbf{B}}_2^- = \begin{pmatrix} 0.151 & 0.295 \\ (0.888) & (1.545) \\ 0.023 & 0.047 \\ (0.183) & (0.518) \end{pmatrix}, \\ \widehat{\mathbf{B}}_3^- &= \begin{pmatrix} 0.228 & -0.245 \\ (1.237) & (-1.337) \\ - & - \end{pmatrix}, \\ \widehat{\mathbf{B}}_{10}^- &= \begin{pmatrix} - & - \\ -0.126 & -0.152 \\ (-0.545) & (-101.590) \end{pmatrix}, \end{aligned}$$

**Table 5** Values of the  $Q(\ell)$  and  $Q_2(\ell)$  test statistics. For  $Q_2(\ell)$  the number of bootstrap replicates is set at  $B = 500$ , and the number of simulations is 1000

Test	Series	Lag $\ell$									
		1	2	3	4	5	6	7	8	9	10
$Q(\ell)$	$Y_{1t}$	-1.25	-0.01	-0.14	0.30	1.48	0.26	-0.65	-0.35	0.82	1.66
	$Y_{2t}$	6.66*	7.41*	6.90*	5.57*	4.76*	4.54*	2.66*	1.14	3.03*	3.14*
$Q_2(\ell)$	$(Y_{1t}, Y_{2t})$	53.57 (0.00)	55.71 (0.00)	47.58 (0.00)	31.16 (0.00)	23.78 (0.01)	22.51 (0.01)	7.15 (0.20)	1.31 (0.80)	9.34 (0.12)	12.70 (0.06)

\* significantly different from zero at the 5% nominal level

**Table 6** ML parameter estimates of an asMA(1) model, their t-values (in parentheses), residual variance  $\hat{\sigma}_\varepsilon^2$ , AIC, and results of the Wald-type test statistic for the first differences of the logarithms of six univariate time series

Series	$\hat{\mu}$	$\hat{b}^+$	$\hat{b}^-$	$\hat{\sigma}_\varepsilon^2 \times 10^5$	AIC	$\widehat{W}_T$
<i>CSSS hot rolled</i>						
$Y_{1t}^{BLS}$	0.000 (0.919)	0.044 (0.202)	0.069 (0.080)	4.05	-10.06	$6.75 \times 10^{-4}$
$Y_{1t}^{NBER}$	0.002 (2.114)	-0.101 (-1.207)	0.339 (1.473)	5.59	-9.74	2.70
<i>Tin plate</i>						
$Y_{2t}^{BLS}$	0.000 (0.515)	-0.006 (-0.001)	-0.814 (-2.301)	3.23	-10.29	$1.87 \times 10^{-2}$
$Y_{2t}^{NBER}$	0.000 (0.361)	0.177 (0.918)	-0.020 (-0.056)	4.16	-10.04	$1.66 \times 10^{-1}$
<i>Regular gasoline</i>						
$Y_{3t}^{BLS}$	-0.009 (-3.003)	0.425 (2.606)	-0.537 (-3.764)	37.4	-7.84	13.97*
$Y_{3t}^{NBER}$	-0.001 (-0.835)	0.170 (1.109)	0.230 (1.977)	2.11	-10.72	$6.95 \times 10^{-3}$

\*  $p$ -value < 0.05

$$\widehat{\mathbf{B}}_1^+ = \begin{pmatrix} 0.615 & -0.654 \\ (4.301) & (-2.024) \\ 0.316 & 1.120 \\ (1.799) & (5.980) \end{pmatrix}, \widehat{\mathbf{B}}_2^+ = \begin{pmatrix} 0.256 & -0.908 \\ (1.497) & (-3.475) \\ 0.310 & 0.061 \\ (2.276) & (0.420) \end{pmatrix},$$

$$\widehat{\mathbf{B}}_3^+ = \begin{pmatrix} -0.265 & 0.009 \\ (-2.114) & (0.038) \\ - & - \end{pmatrix},$$

$$\widehat{\mathbf{B}}_{10}^+ = \begin{pmatrix} - & - \\ -0.004 & 0.081 \\ (-0.031) & (0.570) \end{pmatrix}, \widehat{\Sigma}_\varepsilon = \begin{pmatrix} 4.27 & 0.13 \\ 0.13 & 2.97 \end{pmatrix} \times 10^{-5},$$

with  $t$ -values in parentheses. The Wald-type test statistic is  $\widehat{W}_T = 142.60$  and  $AIC = -20.04$ .

The Wald-type test statistic is asymptotically distributed as  $\chi_{12}^2$  with a  $p$ -value of zero, and hence the  $\mathbb{H}_0$  of symmetry is rejected at the 5% nominal level. We see that for  $Y_{1t}$  the  $t$ -values of the parameter estimates  $\widehat{b}_{12,1}^-$ ,  $\widehat{b}_{11,1}^+$ ,  $\widehat{b}_{12,1}^+$ ,  $\widehat{b}_{12,2}^+$ , and  $\widehat{b}_{11,3}^+$  are all significant. It implies that for this series there is a strong interrelationship between the innovations  $\varepsilon_{2,t-1}^-$ ,  $\varepsilon_{1,t-1}^+$ ,  $\varepsilon_{2,t-1}^+$ ,  $\varepsilon_{2,t-2}^+$  and  $\varepsilon_{1,t-3}^+$  of both time series  $Y_{1t}$  and  $Y_{2t}$ . On the other hand, for  $Y_{2t}$  we see significant parameter estimates  $\widehat{b}_{22,1}^-$ ,  $\widehat{b}_{22,10}^-$ ,  $\widehat{b}_{22,1}^+$ , and  $\widehat{b}_{21,2}^+$ . In this case the dynamic structure depends on the innovations  $\varepsilon_{2,t-1}^-$ ,  $\varepsilon_{2,t-10}^-$ ,  $\varepsilon_{2,t-1}^+$ , and  $\varepsilon_{1,t-2}^+$ . The fitted asVMA model is in agreement with common expectations, i.e., US inflation has a negative, and statistically significant, asymmetric effect on growth rates in US real GDP.

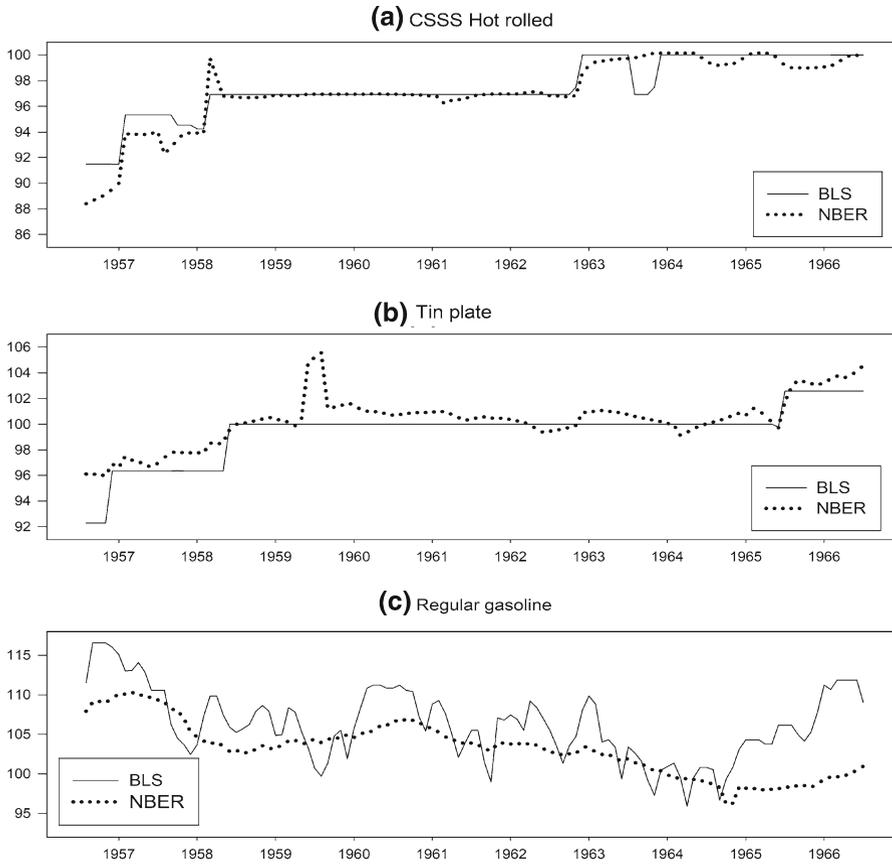


Fig. 2 Six industrial production series for the period 1957:01–1966:12

## 5.2 Production price indices

As a second illustration, we reconsider the second application of Wecker (1981, Sect. 4). He fitted univariate asMA(1) models to six series of the first differences of logarithms of monthly industrial price series ( $T = 119$ ), i.e., production price indices. Three series were obtained from the National Bureau of Labor Statistics (BLS) and three series from the National Bureau of Economic Research (NBER): Carbon Steel, Sheet and Strip – Hot rolled, Tin plate, and Regular gasoline. The BLS series are based on “spot” and quoted prices, while the NBER series are based on contract prices. Figure 2 shows time plots of the original series, initially compiled by Stigler and Kindahl (1970); see [www.nber.org/chapters/c3321.pdf](http://www.nber.org/chapters/c3321.pdf).

Stigler and Kindahl (1970) noted that there are no systematic trend differences between the indices for the period 1957–1961, while for the period 1962–1966 the BLS-based indices rose about 0.7 percent a year relative to the NBER-based indices. They argue that this phenomenon is due to an asymmetric inertia (delay) in industrial

prices movements. Price quotations are not revised immediately when, for instance, market conditions, and transaction prices, change. The asymmetry comes from the prudence of sellers to be slow in authenticating price reduction and prompt in authenticating price increases. This behavior is best represented by spot or quoted prices, and therefore by the BLS-based indices. For a buyer it is measured by the price of a contract, and therefore by the NBER-based indices. Given these observations, the BLS-based production indices are likely to show more asymmetries in sign-based persistence of the innovations than the NBER-based indices.

Below, we first investigate the presence of asymmetries by estimating and testing univariate asMA models for each of the six production series separately. Next, we capture asymmetries for the three BLS-based (NBER-based) indices jointly as a group by fitting asVMA models.

### 5.2.1 asMA models

Wecker (1981) noted that the three BLS-based indices, denoted by  $Y_{it}^{\text{BLS}}$  ( $i = 1, 2, 3$ ), showed statistically significant signs of asymmetry using a likelihood ratio (LR) test statistic of the form  $2 \log(\hat{\sigma}_s^2 / \hat{\sigma}_\varepsilon^2)$  where  $\hat{\sigma}_s^2$  is the estimated standard deviation of the innovations under  $\mathbb{H}_0$  (symmetry) and where  $\hat{\sigma}_\varepsilon^2$  is the estimated standard deviation under  $\mathbb{H}_1$  (asymmetry). No evidence for asymmetry was found in the three NBER-based indices, denoted by  $Y_{it}^{\text{NBER}}$  ( $i = 1, 2, 3$ ). For more information about asymmetric effects, however, it is reasonable to fit asMA(1) models to the univariate time series first. Table 6 summarizes estimation and testing results. The ML parameter estimates are close to those reported by Wecker (1981, Table 3). But note that in almost all cases the estimates of  $b^\pm$  are not statistically different from zero at the 5% nominal level. Moreover, with one exception, the Wald-type test statistic indicates that there is little evidence to reject the null hypothesis of symmetry.

The entries in the first six rows of Table 7 are values of the  $Q(\ell)$  ( $\ell = 1, \dots, 10$ ) test statistic. We see that except for the BLS-CSSS Hot rolled series at lag 8, there is no evidence to reject the null hypothesis  $\mathbb{H}_0: \tilde{\gamma}(\ell) = 0$  at the 5% nominal level. The null hypothesis is rejected, however, for the NBER-CSSS Hot rolled series at lags 1, 2, 6, and 7. Values of  $Q_3(\ell)$  and their corresponding  $p$ -values are reported in rows 7 and 8 of Table 7. For the NBER-CSSS Hot rolled series the test rejects the null hypothesis in (15) at lag 3, using a 5% nominal level. By contrast, for all remaining lags,  $Q_3(\ell)$  gives a strong indication not to reject the hypothesis of symmetry, on the basis of the  $p$ -values. This information may be used to drop insignificant parameter values from the final model specification and, as a result, improves model interpretation.

### 5.2.2 asVMA models

We employ AIC as a model selection criterion but with the additional condition that the final selected model should be parsimonious. For a fully specified asVMA model of dimension  $m = 3$  it is often hard to figure out what is going on. In that case, it makes sense to drop insignificant parameters. Given this objective, we obtain the following

**Table 7** Values of the  $Q(\ell)$  and  $Q_3(\ell)$  test statistics

Test	Series	Lag $\ell$										
		1	2	3	4	5	6	7	8	9	10	
$Q(\ell)$	<i>CSSS hot rolled</i>											
	$Y_{1t}^{BLS}$	-0.47	-0.65	0.42	4.61	-0.25	-0.75	-0.75	-0.75	-7.64*	-0.92	-0.76
	$Y_{1t}^{NBER}$	-7.19*	-6.83*	-0.79	-0.60	-0.59	-4.53*	3.18*	3.18*	-0.33	0.18	-0.47
	<i>Tin plate</i>											
	$Y_{2t}^{BLS}$	-0.88	-0.95	-0.96	-0.97	-0.97	-0.98	-0.98	-0.98	-0.99	-0.99	-0.89
	$Y_{2t}^{NBER}$	1.84	1.77	-14.18*	-0.72	-1.67	-0.22	-0.37	-0.37	-2.27	-1.31	-0.92
	<i>Regular gasoline</i>											
	$Y_{3t}^{BLS}$	1.59	1.39	-1.31	-1.30	0.94	0.63	0.77	0.77	-0.94	-1.78	-1.11
	$Y_{3t}^{NBER}$	-1.07	3.19*	-0.95	0.90	0.25	-0.55	0.72	0.72	0.54	0.26	1.55
	$Q_3(\ell)$	<i>CSSS hot rolled, Tin plate, Regular gasoline</i>										
BLS		5.91 (0.71)	108.94 (0.08)	20.26 (0.36)	25.50 (0.30)	4.48 (0.77)	5.10 (0.73)	14.93 (0.43)	14.93 (0.43)	71.35 (0.13)	10.36 (0.55)	25.08 (0.38)
NBER		59.44 (0.15)	49.71 (0.17)	251.15 (0.03)	1.32 (0.92)	5.11 (0.69)	21.85 (0.30)	11.39 (0.47)	11.39 (0.47)	10.00 (0.51)	15.81 (0.38)	17.27 (0.35)

For  $Q_3(\ell)$  the number of bootstrap replicates is set at  $B = 500$ , and the number of simulations is 1000

\* significantly different from zero at the 5% nominal level

estimation results for the BLS-based indices

$$\begin{aligned} \widehat{\boldsymbol{\mu}} &= \begin{pmatrix} 0.002 \\ (1.000) \\ 0.000 \\ (1.000) \\ -0.010 \\ (-8.25) \end{pmatrix}, \widehat{\mathbf{B}}_1^+ = \begin{pmatrix} - & - & -0.007 \\ & & (-0.056) \\ 0.032 & -0.147 & -0.006 \\ (0.156) & (-0.344) & (-0.138) \\ 0.122 & -0.107 & 0.453 \\ (0.314) & (-0.216) & (11.032) \end{pmatrix}, \\ \widehat{\mathbf{B}}_1^- &= \begin{pmatrix} - & - & 0.081 \\ & & (0.375) \\ 0.156 & -0.400 & -0.064 \\ (0.203) & (-4.130) & (-3.564) \\ 0.229 & -0.417 & -0.593 \\ (1.072) & (-1.195) & (-456.000) \end{pmatrix}, \\ \widehat{\mathbf{B}}_2^+ &= \begin{pmatrix} 0.063 & 0.029 & - \\ (0.141) & (0.038) & \\ - & - & - \\ - & - & - \end{pmatrix}, \widehat{\mathbf{B}}_2^- = \begin{pmatrix} 0.238 & 0.106 & - \\ (0.351) & (0.044) & \\ - & - & - \\ - & - & - \end{pmatrix}, \\ \widehat{\mathbf{B}}_4^+ &= \begin{pmatrix} -0.046 & - & - \\ (-0.102) & - & - \\ - & - & - \\ - & - & - \end{pmatrix}, \widehat{\mathbf{B}}_4^- = \begin{pmatrix} -0.325 & - & - \\ (-0.485) & - & - \\ - & - & - \\ - & - & - \end{pmatrix}, \widehat{\boldsymbol{\Sigma}}_\varepsilon = \begin{pmatrix} 4.02 & -0.25 & 0.81 \\ -0.25 & 1.98 & -0.74 \\ 0.81 & -0.74 & 37.11 \end{pmatrix} \times 10^{-5}, \end{aligned}$$

with  $\widehat{W}_T = 3.82$  and  $\text{AIC} = -27.71$ . The Wald-type test statistic is asymptotically distributed as  $\chi_{10}^2$  with a  $p$ -value of 0.95, and hence the  $\mathbb{H}_0$  of symmetry given by (9) is not rejected at the 5% nominal level. Note that only four parameter estimates  $\widehat{b}_{33,1}^+$ ,  $\widehat{b}_{22,1}^-$ ,  $\widehat{b}_{23,1}^-$ , and  $\widehat{b}_{33,1}^-$  are significantly different from zero.

Similarly, for the three NBER-based indices the “best” asVMA model is given by

$$\begin{aligned} \widehat{\boldsymbol{\mu}} &= \begin{pmatrix} 0.000 \\ (0.600) \\ 0.000 \\ (0.136) \\ -0.001 \\ (-0.600) \end{pmatrix}, \widehat{\mathbf{B}}_1^+ = \begin{pmatrix} - & - & 0.055 \\ & & (0.178) \\ -0.041 & 0.215 & -0.071 \\ (-0.252) & (0.909) & (-0.154) \\ -0.038 & 0.211 & 0.214 \\ (-0.233) & (7.389) & (0.615) \end{pmatrix}, \\ \widehat{\mathbf{B}}_1^- &= \begin{pmatrix} - & - & -0.186 \\ & & (-0.600) \\ -0.287 & -0.041 & 0.154 \\ (-0.590) & (-0.154) & (0.962) \\ 0.190 & -0.071 & 0.250 \\ (12.417) & (-0.312) & (0.698) \end{pmatrix}, \\ \widehat{\mathbf{B}}_2^+ &= \begin{pmatrix} -0.164 & -0.012 & - \\ (-1.028) & (-0.652) & \\ - & - & - \\ - & - & - \end{pmatrix}, \widehat{\mathbf{B}}_2^- = \begin{pmatrix} -0.079 & 0.083 & - \\ (-0.198) & (0.034) & \\ - & - & - \\ - & - & - \end{pmatrix}, \\ \widehat{\mathbf{B}}_6^+ &= \begin{pmatrix} 0.010 & - & - \\ (0.066) & - & - \\ - & - & - \\ - & - & - \end{pmatrix}, \widehat{\mathbf{B}}_6^- = \begin{pmatrix} -0.051 & - & - \\ (-0.116) & - & - \\ - & - & - \\ - & - & - \end{pmatrix}, \widehat{\boldsymbol{\Sigma}}_\varepsilon = \begin{pmatrix} 4.10 & 0.48 & 0.17 \\ 0.48 & 3.97 & -0.21 \\ 0.17 & -0.21 & 2.21 \end{pmatrix} \times 10^{-5}, \end{aligned}$$

with  $\widehat{W}_T = 101.44$  and  $\text{AIC} = -29.84$ .

In this case, the Wald-type test statistic is asymptotically distributed as  $\chi_{10}^2$  with a  $p$ -value of zero, and hence the  $\mathbb{H}_0$  of symmetry given by (9) is rejected at the 5% nominal level. Note that the rejection of  $\mathbb{H}_0$  is most likely due to significant, and positive, values of  $\widehat{b}_{32,1}^+$  and  $\widehat{b}_{31,1}^-$ . Given the size of these parameters, it appears that

positive innovations persist longer than negative innovations. The remaining asVMA parameters are indistinguishable from zero. In summary, our multivariate estimation and testing results cast doubt on Wecker's univariate test results. That is, we only find evidence of asymmetric effects in the three NBER-based indices when tested jointly.

## 6 Summary

In this paper, we introduced the class of asVMA models. The general concept and structure of these models complement the well-known class of univariate asMA models. We derived some basic properties of an asVMA model. We also proposed a multivariate Wald-type test statistic to uncover asymmetric effects in vector time series. Simulation experiments demonstrated reasonable to good power performance of the Wald-type test statistic in finite sample cases. We illustrated the proposed test and estimation procedure by finding evidence of asymmetry in two sets of empirical time series, one set modeled by a two-dimensional asVMA model, and one set by a three-dimensional asVMA model. Given these results, we believe that the asVMA model in conjunction with the proposed test statistic, has great potential to uncover asymmetric phenomena in multivariate time series with more precision than has been possible before.

**Acknowledgements** We would like to thank two anonymous referees for helpful comments and suggestions.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## Appendix: Proofs of results

For ease in presentation, let  $M^{(i)} = \{\boldsymbol{\varepsilon} \in \mathbb{R}^m \mid \mathbf{1}'_m \boldsymbol{\varepsilon} \in \mathbb{R}^{(i)}\}$  ( $i = 1, 2$ ) denote the space  $\mathbb{R}^m$  decomposed into two non-overlapping subspaces, where  $\mathbb{R}^{(1)} = (-\infty, 0)$  and  $\mathbb{R}^{(2)} = [0, \infty)$ . Furthermore, we define the set of matrix functions  $\mathcal{B}: \mathbb{R}^m \rightarrow \mathbb{R}^{m \times m}$  such that

$$\mathcal{B}(\boldsymbol{\varepsilon}_{t-1}) = \mathbf{B}^{(1)} I(\mathbf{1}'_m \boldsymbol{\varepsilon}_{t-1} \in \mathbb{R}^{(1)}) + \mathbf{B}^{(2)} I(\mathbf{1}'_m \boldsymbol{\varepsilon}_{t-1} \in \mathbb{R}^{(2)}),$$

where  $\mathbf{B}^{(1)} \equiv \mathbf{B}^-$  and  $\mathbf{B}^{(2)} \equiv \mathbf{B}^+$ .

The proof of the Proposition requires the following two lemmas.

**Lemma 1** Let  $\boldsymbol{\Sigma}_\varepsilon = [\sigma_{ij}]_{i,j=1}^m = \begin{pmatrix} \sigma_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}'_{12} & \boldsymbol{\Sigma}_{22} \end{pmatrix}$ , where  $\boldsymbol{\Sigma}_{12}$  is an  $1 \times (m-1)$  vector,  $\boldsymbol{\Sigma}_{22}$  is an  $(m-1) \times (m-1)$  matrix, and  $|\boldsymbol{\Sigma}_{22}| > 0$ . Further, let  $f_m(\boldsymbol{\varepsilon})$  denote the density function of  $\{\boldsymbol{\varepsilon}_t, t \in \mathbb{Z}\}$ . Then

- (i)  $\int_{A^{(i)}} \varepsilon_j f_m(\boldsymbol{\varepsilon}) d\boldsymbol{\varepsilon} = \left(\frac{\sigma_{j1}}{\sigma_{11}}\right) \mu^{(i)}, (i = 1, 2; j = 1, \dots, m).$
- (ii)  $\int_{A^{(i)}} \varepsilon_k \varepsilon_j f_m(\boldsymbol{\varepsilon}) d\boldsymbol{\varepsilon} = \left(\frac{\sigma_{k1} \sigma_{j1}}{\sigma_{11}}\right) \left(\frac{\sigma^{(i)}}{\sigma_{11}} - \alpha^{(i)}\right) + \sigma_{kj} \alpha^{(i)}, (i = 1, 2; j, k = 1, \dots, m),$

where

$$A^{(i)} = \{(z_1, \dots, z_m) \in \mathbb{R}^m; z_1 \in \mathbb{R}^{(i)}, (z_2, \dots, z_m) \in \mathbb{R}^{m-1}\},$$

and

$$\alpha^{(i)} = \int_{\mathbb{R}^{(i)}} f_1(u) du, \quad \mu^{(i)} = \int_{\mathbb{R}^{(i)}} u f_1(u) du, \quad \sigma^{(i)} = \int_{\mathbb{R}^{(i)}} u^2 f_1(u) du.$$

**Proof** Let  $U \sim \mathcal{N}(0, \sigma_{11})$  and  $\mathbf{V} \sim \mathcal{N}_{m-1}(\mathbf{0}, \boldsymbol{\Sigma}_{22})$ . Then it is well known from the theory of multivariate statistical analysis that the conditional distribution of  $\mathbf{V}$  given that  $U = u$ , is normal with mean  $\boldsymbol{\Sigma}_{12} \sigma_{11}^{-1} u$  and covariance matrix  $\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{12} \sigma_{11}^{-1} \boldsymbol{\Sigma}'_{12}$ . Denote the corresponding conditional density function by  $g_{m-1}(\cdot|u)$ .

- (i) Since  $f_m(\mathbf{v}) = g_{m-1}(\mathbf{v}|u) f_1(u)$  with  $\mathbf{v} = (v_1, \dots, v_m)'$ , we have for  $j = 1$

$$\begin{aligned} \int_{A^{(i)}} \varepsilon_1 f_m(\boldsymbol{\varepsilon}) d\boldsymbol{\varepsilon} &= \int_{A^{(i)}} u f_1(u) g_{m-1}(\mathbf{v}|u) du d\mathbf{v} \\ &= \int_{\mathbb{R}^{(i)}} u f_1(u) \left( \int_{\mathbb{R}^{m-1}} g_{m-1}(\mathbf{v}|u) d\mathbf{v} \right) du = \mu^{(i)}. \end{aligned}$$

In a similar way, for  $j > 1$ ,

$$\begin{aligned} \int_{A^{(i)}} \varepsilon_j f_m(\boldsymbol{\varepsilon}) d\boldsymbol{\varepsilon} &= \int_{\mathbb{R}^{(i)}} f_1(u) \left( \int_{\mathbb{R}^{m-1}} v_j g_{m-1}(\mathbf{v}|u) d\mathbf{v} \right) du \\ &= \int_{\mathbb{R}^{(i)}} f_1(u) \frac{\sigma_{j1}}{\sigma_{11}} u du = \left(\frac{\sigma_{j1}}{\sigma_{11}}\right) \mu^{(i)}. \end{aligned}$$

- (ii) Clearly, for  $k = j = 1$  the term on the right-hand side becomes  $\sigma^{(i)} - \sigma_{11} \alpha^{(i)} + \sigma_{11} \alpha^{(i)} = \sigma^{(i)}$ . Then, similar to part (i), we have

$$\begin{aligned} \int_{A^{(i)}} \varepsilon_1^2 f_m(\boldsymbol{\varepsilon}) d\boldsymbol{\varepsilon} &= \int_{A^{(i)}} u^2 f_1(u) g_{m-1}(\mathbf{v}|u) du d\mathbf{v} \\ &= \int_{\mathbb{R}^{(i)}} u^2 f_1(u) \left( \int_{\mathbb{R}^{m-1}} g_{m-1}(\mathbf{v}|u) d\mathbf{v} \right) du = \sigma^{(i)}. \end{aligned}$$

For  $k = 1, j \neq 1$ , the term on the right-hand side becomes  $(\sigma_{j1}/\sigma_{11}) \sigma^{(i)}$ . Again, this result follows in a similar way as above, i.e.,

$$\int_{A^{(i)}} \varepsilon_1 \varepsilon_j f_m(\boldsymbol{\varepsilon}) d\boldsymbol{\varepsilon} = \int_{A^{(i)}} u f_1(u) v_j g_{m-1}(\mathbf{v}|u) du d\mathbf{v}$$

$$\begin{aligned}
 &= \int_{\mathbb{R}^{(i)}} u f_1(u) \left( \int_{\mathbb{R}^{m-1}} v_j g_{m-1}(\mathbf{v}|u) d\mathbf{v} \right) du \\
 &= \int_{\mathbb{R}^{(i)}} u f_1(u) \frac{\sigma_{j1}}{\sigma_{11}} u du = \left( \frac{\sigma_{j1}}{\sigma_{11}} \right) \sigma^{(i)}.
 \end{aligned}$$

For  $k \neq 1, j \neq 1$ , we have

$$\begin{aligned}
 &\int_{A^{(i)}} \varepsilon_k \varepsilon_j f_m(\boldsymbol{\varepsilon}) d\boldsymbol{\varepsilon} \\
 &= \int_{A^{(i)}} f_1(u) \left( \int_{\mathbb{R}^{m-1}} v_k v_j g_{m-1}(\mathbf{v}|u) d\mathbf{v} \right) du \\
 &= \int_{\mathbb{R}^{(i)}} f_1(u) \left\{ \int_{\mathbb{R}^{m-1}} \left( (v_k - \frac{\sigma_{k1}}{\sigma_{11}}u)(v_j - \frac{\sigma_{j1}}{\sigma_{11}}u) \right. \right. \\
 &\quad \left. \left. + \frac{\sigma_{k1}}{\sigma_{11}}u v_j + \frac{\sigma_{j1}}{\sigma_{11}}u v_k - u^2 \frac{\sigma_{k1}\sigma_{j1}}{\sigma_{11}^2} \right) g_{m-1}(\mathbf{v}|u) d\mathbf{v} \right\} du \\
 &= \int_{\mathbb{R}^{(i)}} f_1(u) \left( \sigma_{kj} - \frac{\sigma_{k1}\sigma_{j1}}{\sigma_{11}} + \frac{\sigma_{k1}}{\sigma_{11}}u \frac{\sigma_{j1}}{\sigma_{11}} + \frac{\sigma_{j1}}{\sigma_{11}}u \frac{\sigma_{k1}}{\sigma_{11}} - u^2 \frac{\sigma_{k1}\sigma_{j1}}{\sigma_{11}^2} \right) du \\
 &= \left( \sigma_{kj} - \frac{\sigma_{k1}\sigma_{j1}}{\sigma_{11}} \right) \alpha^{(i)} + \frac{\sigma_{k1}\sigma_{j1}}{\sigma_{11}^2} \sigma^{(i)} = \frac{\sigma_{k1}\sigma_{j1}}{\sigma_{11}} \left( \frac{\sigma^{(i)}}{\sigma_{11}} - \alpha^{(i)} \right) + \sigma_{kj} \alpha^{(i)}.
 \end{aligned}$$

□

**Lemma 2** Let  $\mathbf{r}$  and  $\mathbf{s}$  be two  $m$ -dimensional non-random vectors in  $\mathbb{R}^m$ . Then, using Lemma 1, it follows that

- (i)  $\int_{A^{(i)}} \mathbf{r}' \boldsymbol{\varepsilon} f_m(\boldsymbol{\varepsilon}) d\boldsymbol{\varepsilon} = (\mu^{(i)} / \sigma_{11}) \mathbf{r}' \boldsymbol{\Sigma}_{12}^*$ , ( $i = 1, 2$ ), where  $\boldsymbol{\Sigma}_{12}^* = (\sigma_{11}, \boldsymbol{\Sigma}_{12})'$ .
- (ii)  $\int_{A^{(i)}} \mathbf{r}' \boldsymbol{\varepsilon} \boldsymbol{\varepsilon}' \mathbf{s} f_m(\boldsymbol{\varepsilon}) d\boldsymbol{\varepsilon} = \gamma^{(i)} \mathbf{r}' \boldsymbol{\Sigma}_{12}^* \boldsymbol{\Sigma}_{12}^{*'} \mathbf{s} + \alpha^{(i)} \mathbf{r}' \boldsymbol{\Sigma}_e \mathbf{s}$ , ( $i = 1, 2$ ), where

$$\gamma^{(i)} = \frac{\sigma^{(i)}}{\sigma_{11}^2} - \frac{\alpha^{(i)}}{\sigma_{11}}.$$

**Proof** (i) Using result (i) of Lemma 1, we have

$$\begin{aligned}
 \int_{A^{(i)}} \mathbf{r}' \boldsymbol{\varepsilon} f_m(\boldsymbol{\varepsilon}) d\boldsymbol{\varepsilon} &= \int_{A^{(i)}} \sum_{j=1}^m r_j \varepsilon_j f_m(\boldsymbol{\varepsilon}) d\boldsymbol{\varepsilon} \\
 &= \sum_{j=1}^m r_j \int_{A^{(i)}} \varepsilon_j f_m(\boldsymbol{\varepsilon}) d\boldsymbol{\varepsilon} = \frac{\mu^{(i)}}{\sigma_{11}} \mathbf{r}' \boldsymbol{\Sigma}_{12}^*.
 \end{aligned}$$

(ii) Using result (ii) of Lemma 1, we have

$$\int_{A^{(i)}} \mathbf{r}' \boldsymbol{\varepsilon} \boldsymbol{\varepsilon}' \mathbf{s} f_m(\boldsymbol{\varepsilon}) d\boldsymbol{\varepsilon} = \int_{A^{(i)}} \left( \sum_{j=1}^m r_j \varepsilon_j \right) \left( \sum_{k=1}^m s_k \varepsilon_k \right) f_m(\boldsymbol{\varepsilon}) d\boldsymbol{\varepsilon}$$

$$\begin{aligned}
 &= \sum_{j=1}^m \sum_{k=1}^m r_j s_k (\sigma_{kj} \sigma_{j1} \gamma^{(i)} + \sigma_{kj} \alpha^{(i)}) \\
 &= \gamma^{(i)} \mathbf{r}' \boldsymbol{\Sigma}_{12}^* \boldsymbol{\Sigma}_{12}^{*'} \mathbf{s} + \alpha^{(i)} \mathbf{r}' \boldsymbol{\Sigma}_\varepsilon \mathbf{s}.
 \end{aligned}$$

□

**Proof of Proposition** To obtain the required moments, we introduce an  $m \times m$  orthogonal matrix  $\mathbf{Q}$  with first column given by  $\mathbf{1}_m$ . Applying the nonsingular linear transformation  $\mathbf{Q}'\boldsymbol{\varepsilon}$  to  $\boldsymbol{\varepsilon}$  with Jacobian  $|\mathbf{Q}| \neq 0$ , we have

$$\mathbf{Q}'\boldsymbol{\varepsilon} \in A^{(i)} \iff \boldsymbol{\varepsilon} \in M^{(i)}, \quad (i = 1, 2).$$

Let  $\varphi(\boldsymbol{\varepsilon})$  be the density function of  $\mathcal{N}_m(\mathbf{0}, \boldsymbol{\Sigma}_\varepsilon)$ , and  $\psi(\mathbf{z})$  the density function of  $\mathcal{N}_m(\mathbf{0}, \mathbf{Q}'\boldsymbol{\Sigma}_\varepsilon\mathbf{Q})$ . Then there exists a bounded continuous function  $g(\cdot)$  such that

$$\int_{M^{(i)}} g(\boldsymbol{\varepsilon})\varphi(\boldsymbol{\varepsilon})d\boldsymbol{\varepsilon} = \int_{\xi(M^{(i)})} g(\xi^{-1}(\boldsymbol{\varepsilon}))\varphi(\xi^{-1}(\boldsymbol{\varepsilon}))d\boldsymbol{\varepsilon} = \int_{A^{(i)}} g(\mathbf{Q}\mathbf{z})\psi(\mathbf{z})d\mathbf{z}, \quad (i = 1, 2),$$

with  $\xi: \boldsymbol{\varepsilon} \rightarrow \mathbf{Q}'\boldsymbol{\varepsilon}$ . Hence, the first moment of  $\{\mathbf{Y}_t, t \in \mathbb{Z}\}$  is given by

$$\begin{aligned}
 \mathbb{E}(\mathbf{Y}_t) &= \mathbb{E}(\boldsymbol{\varepsilon}_t) + \mathbb{E}[\mathbf{B}(\boldsymbol{\varepsilon}_{t-1})\boldsymbol{\varepsilon}_{t-1}] \\
 &= \mathbb{E}(\boldsymbol{\varepsilon}_t) + \sum_{i=1}^2 \int_{M^{(i)}} \mathbf{B}^{(i)} \boldsymbol{\varepsilon} \varphi(\boldsymbol{\varepsilon}) d\boldsymbol{\varepsilon} = \sum_{i=1}^2 \int_{A^{(i)}} \mathbf{B}^{(i)} \mathbf{Q}\mathbf{z} \psi(\mathbf{z}) d\mathbf{z}.
 \end{aligned}$$

Note that the covariance matrix  $\mathbf{Q}'\boldsymbol{\Sigma}_\varepsilon\mathbf{Q}$  has  $\mathbf{Q}'\boldsymbol{\Sigma}_\varepsilon\mathbf{1}_m$  as first column. Then, using result (i) of Lemma 2, we have

$$\begin{aligned}
 \mathbb{E}(\mathbf{Y}_t) &= \sum_{i=1}^2 \frac{\mu^{(i)}}{\mathbf{1}'_m \boldsymbol{\Sigma}_\varepsilon \mathbf{1}_m} \mathbf{B}^{(i)} \mathbf{Q} \mathbf{Q}' \boldsymbol{\Sigma}_\varepsilon \mathbf{1}_m = \frac{1}{\mathbf{1}'_m \boldsymbol{\Sigma}_\varepsilon \mathbf{1}_m} \sum_{i=1}^2 \mu^{(i)} \mathbf{B}^{(i)} \boldsymbol{\Sigma}_\varepsilon \mathbf{1}_m \\
 &= \frac{1}{\mathbf{1}'_m \boldsymbol{\Sigma}_\varepsilon \mathbf{1}_m} \mu^{(2)} (\mathbf{B}^{(1)} - \mathbf{B}^{(2)}) \boldsymbol{\Sigma}_\varepsilon \mathbf{1}_m = (2\pi \mathbf{1}'_m \boldsymbol{\Sigma}_\varepsilon \mathbf{1}_m)^{-1/2} (\mathbf{B}^{(1)} - \mathbf{B}^{(2)}) \boldsymbol{\Sigma}_\varepsilon \mathbf{1}_m,
 \end{aligned}$$

since  $-\mu^{(1)} = \mu^{(2)} = (\mathbf{1}'_m \boldsymbol{\Sigma}_\varepsilon \mathbf{1}_m / 2\pi)^{1/2}$ .

The variance of  $\{\mathbf{Y}_t, t \in \mathbb{Z}\}$  is given by  $\text{Var}(\mathbf{Y}_t) = \mathbb{E}(\mathbf{Y}_t \mathbf{Y}'_t) - \mathbb{E}(\mathbf{Y}_t) \mathbb{E}(\mathbf{Y}'_t)$ . Using result (ii) of Lemma 2, the first term on the right-hand side can be written as

$$\begin{aligned}
 \mathbb{E}(\mathbf{Y}_t \mathbf{Y}'_t) &= \boldsymbol{\Sigma}_\varepsilon + \sum_{i=1}^2 \int_{M^{(i)}} \mathbf{B}^{(i)} \boldsymbol{\varepsilon} \boldsymbol{\varepsilon}' (\mathbf{B}^{(i)})' \varphi(\boldsymbol{\varepsilon}) d\boldsymbol{\varepsilon} = \boldsymbol{\Sigma}_\varepsilon \\
 &\quad + \sum_{i=1}^2 \int_{A^{(i)}} \mathbf{B}^{(i)} \mathbf{Q} \mathbf{z} \mathbf{z}' \mathbf{Q}' (\mathbf{B}^{(i)})' \psi(\mathbf{z}) d\mathbf{z}
 \end{aligned}$$

$$\begin{aligned} &= \Sigma_\varepsilon + \sum_{i=1}^2 \gamma^{(i)} \mathbf{B}^{(i)} \Sigma_\varepsilon \mathbf{1}_m \mathbf{1}'_m \Sigma'_\varepsilon (\mathbf{B}^{(i)})' + \sum_{i=1}^2 \alpha^{(i)} \mathbf{B}^{(i)} \Sigma_\varepsilon (\mathbf{B}^{(i)})' \\ &= \Sigma_\varepsilon + \frac{1}{2} (\mathbf{B}^{(1)} \Sigma_\varepsilon (\mathbf{B}^{(1)})' + \mathbf{B}^{(2)} \Sigma_\varepsilon (\mathbf{B}^{(2)})'), \end{aligned}$$

since  $\alpha^{(1)} = \alpha^{(2)} = 1/2$ , and

$$\gamma^{(i)} = \frac{\sigma^{(i)}}{\sigma_{11}^2} - \frac{\alpha^{(i)}}{\sigma_{11}} = \frac{\mathbf{1}'_m \Sigma_\varepsilon \mathbf{1}_m}{2} \frac{1}{(\mathbf{1}'_m \Sigma_\varepsilon \mathbf{1}_m)^2} - \frac{1/2}{\mathbf{1}'_m \Sigma_\varepsilon \mathbf{1}_m} = 0, \quad (i = 1, 2).$$

The cross-covariance matrix at lag  $\ell = 1$  is given by

$$\begin{aligned} \Gamma(1) &= \text{Cov}(\mathbf{Y}_t, \mathbf{Y}_{t-1}) = \text{Cov}(\boldsymbol{\varepsilon}_t, \boldsymbol{\varepsilon}_{t-1}) + \text{Cov}(\boldsymbol{\varepsilon}_t, \mathcal{B}(\boldsymbol{\varepsilon}_{t-2})\boldsymbol{\varepsilon}_{t-2}) \\ &\quad + \text{Cov}(\mathcal{B}(\boldsymbol{\varepsilon}_{t-1})\boldsymbol{\varepsilon}_{t-1}, \boldsymbol{\varepsilon}_{t-1}) + \text{Cov}(\mathcal{B}(\boldsymbol{\varepsilon}_{t-1})\boldsymbol{\varepsilon}_{t-1}, \mathcal{B}(\boldsymbol{\varepsilon}_{t-2})\boldsymbol{\varepsilon}_{t-2}) \\ &= \text{Cov}(\mathcal{B}(\boldsymbol{\varepsilon}_{t-1})\boldsymbol{\varepsilon}_{t-1}, \boldsymbol{\varepsilon}_{t-1}) \\ &= \mathbb{E}(\mathcal{B}(\boldsymbol{\varepsilon}_{t-1})\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}'_{t-1}) - \mathbb{E}(\mathcal{B}(\boldsymbol{\varepsilon}_{t-1})\boldsymbol{\varepsilon}_{t-1})(\mathbb{E}(\boldsymbol{\varepsilon}_{t-1}))' \\ &= \mathbb{E}(\mathcal{B}(\boldsymbol{\varepsilon}_{t-1})\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}'_{t-1}). \end{aligned}$$

Next, using result (ii) of Lemma 2, we have

$$\begin{aligned} \Gamma(1) &= \mathbb{E}(\mathcal{B}(\boldsymbol{\varepsilon}_{t-1})\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}'_{t-1}) = \sum_{i=1}^2 \int_{M^{(i)}} \mathbf{B}^{(i)} \boldsymbol{\varepsilon} \boldsymbol{\varepsilon}' \varphi(\boldsymbol{\varepsilon}) d\boldsymbol{\varepsilon} \\ &= \sum_{i=1}^2 \int_{A^{(i)}} \mathbf{B}^{(i)} \mathbf{Q} \mathbf{z} \mathbf{z}' \mathbf{Q}' \psi(\mathbf{z}) d\mathbf{z} \\ &= \sum_{i=1}^2 \gamma^{(i)} \mathbf{B}^{(i)} \mathbf{Q} \mathbf{Q}' \Sigma_\varepsilon \mathbf{1}_m \mathbf{1}'_m \Sigma'_\varepsilon \mathbf{Q} \mathbf{Q}' + \sum_{i=1}^2 \alpha^{(i)} \mathbf{B}^{(i)} \mathbf{Q} \mathbf{Q}' \Sigma_\varepsilon \mathbf{Q} \mathbf{Q}' \\ &= \left( \sum_{i=1}^2 \gamma^{(i)} \mathbf{B}^{(i)} \right) \Sigma_\varepsilon \mathbf{1}_m \mathbf{1}'_m \Sigma_\varepsilon + \left( \sum_{i=1}^2 \alpha^{(i)} \mathbf{B}^{(i)} \right) \Sigma_\varepsilon \\ &= \frac{1}{2} (\mathbf{B}^{(1)} + \mathbf{B}^{(2)}) \Sigma_\varepsilon, \end{aligned}$$

since  $\gamma^{(i)} = 0$  and  $\alpha^{(i)} = 1/2$  ( $i = 1, 2$ ). When  $\ell > 1$ , it is easy to see that  $\Gamma(\ell) = \mathbf{0}_{m \times m}$ . This completes the proof of the Proposition.  $\square$

### References

Atanasova C (2003) Credit market imperfections and business cycle dynamics: a nonlinear approach. *Stud Nonlinear Dyn Econom* 7(4): Article 5  
 Brännäs K, De Gooijer JG (1994) Autoregressive-asymmetric moving average models for business cycle data. *J Forecast* 13:529–44

- Brännäs K, De Gooijer JG (2004) Asymmetries in conditional mean and variance: modelling stock returns by asMA( $q$ )-asQGARCH. *J Forecast* 23:155–71
- Brännäs K, de Luna X (1998) Generalized method of moment and indirect estimation of the ARasMA model. *Computat. Stat.* 13:485–94
- Brännäs K, Ohlsson H (1999) Asymmetric time series and temporal aggregation. *Rev Econ Stat* 81:341–44
- Brännäs K, De Gooijer JG, Teräsvirta T (1998) Testing linearity against nonlinear moving average models. *Commun Stat Theory Methods* 27:2025–35
- De Gooijer JG, Brännäs K (1995) Invertibility of nonlinear time series models. *Commun Stat Theory Methods* 24:2701–14
- Elwood SK (1998) Is the persistence of shocks to output asymmetric? *J Monet Econ* 41:411–26
- Gonzalo J, Martínez O (2006) Large shocks versus small shocks. (Or does size matter? May be so). *J Econom* 135:311–47
- Guay A, Scaillet O (2003) Indirect inference, nuisance parameter, and threshold moving average models. *J Bus Econ Stat* 21:122–32
- Keating JW (2000) Macroeconomic modeling with asymmetric vector autoregressions. *J Macroecon* 22:1–28
- Koutmos G (1999) Asymmetric index stock returns: evidence from the G-7. *Appl Econ Lett* 6:817–20
- Nebeling T, Salish N (2017) On testing shock induced asymmetries in time series. [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2672999](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2672999)
- Niglio M, Vitale CD (2013) Vector threshold moving average models: model specification and invertibility. In: Torelli N et al (eds) *Advances in theoretical and applied statistics*. Springer-Verlag, New York, pp 87–98
- Patton A, Politis DN, White H (2007) Correction to automatic block-length selection for the dependent bootstrap. *Econom Theory* 29:372–75
- Politis DN, White H (2004) Automatic block-length selection for the dependent bootstrap. *Econom Rev* 23:53–70
- Reinsel GC, Basu S, Yap SF (1992) Maximum likelihood estimators in the multivariate autoregressive moving average model from a generalized least squares viewpoint. *J Time Ser Anal* 13:133–45
- Sáfadi T, Morettin PA (1998) A Bayesian analysis of asymmetric time series. *Br J Probab Stat* 12:1–15
- Stigler GJ, Kindahl JK (1970) *The behavior of industrial prices*. National Bureau of Economic Research, New York
- Taştan H (2017) Simulation-based estimation of threshold moving average models with contemporaneous shock asymmetry and an application to Turkish business cycles. *Commun Stat Simul Comput* 46:3870–91
- Wecker WE (1981) Asymmetric time series. *J Am Stat Assoc* 76:16–21 Corrigenda: p 954
- Weisse CL (1999) The asymmetric effects of monetary policy: a nonlinear vector autoregression approach. *J Money Credit Bank* 31:85–108
- Welsh AK, Jernigan RW (1983) A statistic to identify asymmetric time series. In: *American Statistical Association: Proceedings of the business and economic section*, pp 390–95
- White H (2001) *Asymptotic theory for econometricians*, Rev edn. Academic Press, London

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.