Enforcing solvability of a nonlinear matrix equation and estimation of multivariate ARMA time series

Tobias Brüll[*¶] and Federico Poloni[†] and
Giacomo Sbrana[‡] and Christian Schröder[§¶]

July 18, 2013

Abstract

The matrix equation $X + AX^{-1}A^T = B$, arising in parameter estimation of certain time series models, is solvable only for certain values of the matrices A, B. We present a numerical method to modify A, B in order to make the matrix equation solvable. Since solvability depends on the location of the eigenvalues of the palindromic matrix polynomial $\lambda^2 A + \lambda B + A^T$, our method works by moving those eigenvalues to specified locations using first order spectral perturbation theory. The method is heuristic but works in practice, as is supported by several compelling numerical examples. These examples arise from parameter estimation of a common time series model, the multivariate ARMA(1,1).

Keywords: regularization, palindromic eigenvalue problem, first order eigenvalue perturbation theory, parameter estimation, VARMA(1,1) model
MSC2000 classification: 65F30, 15A24, 62M10, 91B84, 93E10

1 Introduction

Time series models play an important role in applications, for instance in econometrics [7, 22]. In order to return meaningful results, the parameters of a model need to be fit to the observed data, or *estimated*. We discuss a parameter estimation procedure, one of whose steps consists of solving the nonlinear matrix equation

$$X + AX^{-1}A^T = B, \tag{1}$$

where $A, B \in \mathbb{R}^{n,n}$ with $B > 0$ (i.e., B is Hermitian and positive definite). We are interested in making sure that there is a solution $X \in \mathbb{R}^{n,n}$, $X > 0$. It is known (e.g., [8]) that such a solution exists if and only if the matrix Laurent polynomial

$$Q(\lambda) = \lambda A + B + \lambda^{-1} A^T$$

is regular (i.e., the matrix $Q(\lambda)$ is nonsingular for at least one value of $\lambda \in \mathbb{C}$) and $Q(\lambda) \geq 0$ (i.e., $Q(\lambda)$ is Hermitian and positive semidefinite) for each complex value λ on the unit circle.

[*]CST – Computer Simulation Technology AG, Bad Nauheimer Str. 19, 64289 Darmstadt, Germany. Tobias.Bruell@cst.com.

[†]Dipartimento di Informatica, Università di Pisa, Largo Pontecorvo 3, 56127 Pisa, Italy. fpoloni@di.unipi.it. Supported in part by the A. von Humboldt foundation and by INDAM (Istituto Nazionale di Alta Matematica).

[‡]Rouen Business School, 1 Rue du Maréchal Juin, 76130 Mont-Saint-Aignan, France. gsb@rouenbs.fr.

[§]Institut für Mathematik, Technische Universität Berlin, Str. des 17. Juni 136, 10623 Berlin, Germany. schroed@math.tu-berlin.de.

[¶]Research supported by *Deutsche Forschungsgemeinschaft*, via the DFG Research Center MATHEON in Berlin

Moreover, a solution X with $\rho(X^{-1}A^T) < 1$ (as it is needed in applications; where $\rho(\cdot)$ denotes the spectral radius) exists if and only if $Q(\lambda) > 0$ for each unimodular λ. Assuming positive definiteness of $Q(\lambda)$ for at least one such λ, the last condition is equivalent to stating that Q has no generalized eigenvalues on the unit circle. We note that Q is palindromic, i.e., $Q(\lambda^{-1}) = Q(\lambda)^T$. For these Laurent polynomials it is known that generalized eigenvalues on the unit circle are a generic possibility (i.e., they do not occur on a measure-zero subset of the parameters only) [23, 28].

In practice, often the coefficients A and B are affected by errors, e.g., because they come out of data measurements, or their determination involves some form of linearization, truncation, or other such simplifications. Then it may well be the case that the original intended matrix equation admits a solution, whereas the perturbed one – which is available in practice – does not.

In this paper, we propose an algorithm to compute perturbations $\tilde{A} = A + E$, $\tilde{B} = B + F$, with $\|E\|$ and $\|F\|$ small, such that Equation (1) (with A, B replaced by $\tilde{A}, \tilde{B}$) is guaranteed to be solvable. This is achieved by removing all generalized eigenvalues of $\tilde{Q}(\lambda) = \lambda\tilde{A} + \tilde{B} + \lambda^{-1}\tilde{A}^T$ from the unit circle. The presented method is a modification of the method in [5]. Instead of moving eigenvalues of palindromic matrix Laurent polynomials off the unit circle, there eigenvalues of symmetric-skew-symmetric pencils are moved off the imaginary axis. The motivation there is the passivation of linear time invariant dynamical systems, and the method is based upon earlier work in this direction [1, 4, 12, 29, 30, 31]. Other related methods include [14, 13] where pseudospectral methods are used and [2] where negative imaginariness of descriptor systems is enforced. In contrast to our method, those in [1, 13] are guaranteed to find the smallest perturbation, but only for merging two eigenvalues at a time, and they do not allow restricting the perturbation to a linear space. We refer the reader to [5] for further discussion of this issue in a related case.

Let us also mention that the matrix equation (1) is related to discrete algebraic Riccati equations (DARE) as follows [8, Proposition 7.1]: if X solves (1) (with $B = I$) then it also solves the DARE $X = AXA^T + AA^T - AX(-I + X)^{-1}XA^T$.

The idea of our algorithm is to take a set of matrix pairs (E_i, F_i), $i = 1, 2, \ldots, m$, and to compute the approximate location of the generalized eigenvalues of $\tilde{Q}$, with

$$\tilde{A} = A + \sum_{i=1}^{m} E_i\delta_i, \quad \tilde{B} = B + \sum_{i=1}^{m} F_i\delta_i,$$

using first-order eigenvalue perturbation results. Such approximate locations depend linearly on the scalar weights δ_i. Hence, by solving a linear system of equations, we can determine a choice of the δ_i, $i = 1, 2, \ldots, m$, that moves the generalized eigenvalues closer to a specified location. Using several steps of this procedure, the eigenvalues can be moved in sufficiently small steps to a different location. In particular, we can move the eigenvalues on the unit circle in directions that make them coalesce into pairs and then leave the circle.

Our motivation for enforcing solvability of (1) came from a practical problem in econometrics. A task often encountered there is the parameter estimation of models for economic time series. In particular, we focus on the multivariate *autoregressive model with moving average*; in short *multivariate/vector ARMA(1,1)* model, or even shorter *VARMA(1,1)* model [22]. Given the parameters $\Phi, \Theta \in \mathbb{R}^{d,d}$, $c \in \mathbb{R}^d$ and randomly drawn noise vectors $\hat{u}_t \in \mathbb{R}^d$, this process produces a vector sequence $(\hat{x}_t)_{t=1,2,\ldots}$ in $\mathbb{R}^d$ by

$$\hat{x}_t = \Phi\hat{x}_{t-1} + c + \hat{u}_t - \Theta\hat{u}_{t-1}, \quad t = 2, 3, \ldots.$$

The task is to recover (i.e., estimate) the parameters Φ, Θ, c from an observed finite subsequence $(\hat{x}_t)_{t=1,\ldots,N}$. As an alternative to the common maximum likelihood estimation procedure [22], or in order to produce good starting values for the same procedure, we propose a moment-based estimator (also known as a method of moments), i.e., an estimator which only uses moments of the time series (here, mean and autocorrelations) to estimate the parameters.

Moment-based estimators exist in the econometrics literature for univariate ARMA processes [20]. An extension to multivariate models, which requires more advanced linear algebra techniques, was suggested in [26] for a specific econometric model, the multivariate GARCH(1,1) model. We describe it here for the more general VARMA(1,1) model.

The basic idea of the procedure consists of computing the finite-sample moments

$$\hat{\mu} := \frac{1}{N}\sum_{t=1}^{N}\hat{x}_t, \quad \hat{M}_k := \frac{1}{N-k}\sum_{t=1}^{N-k}(\hat{x}_{t+k}-\hat{\mu})(\hat{x}_t-\hat{\mu})^T, \tag{2}$$

(which converge to the true mean and autocovariance matrices for $N \to \infty$) and then fitting the parameters to obtain a model that matches the observed moments as closely as possible.

In particular, our moment-based estimator requires the solution of a matrix equation of the form (1). Unfortunately, the finite-sample moments (2) converge rather slowly to the true asymptotic ones, i.e., substantial deviations are not unlikely. Therefore, one may encounter the situation described above, where the matrix equation at hand admits no solution that satisfies all the required assumptions. The regularization technique presented in this paper can then be used to obtain solutions in cases in which the estimator would fail otherwise; the robustness of the resulting method is greatly increased.

This paper is structured as follows: first the solvability enforcement method is developed in sections 2–4. More precisely, Section 2 introduces spectral plots, the main tool to understand how the enforcement works. Section 3 collects a few results on perturbation theory of unimodular eigenvalues of palindromic Laurent polynomials. All parts are then combined to the enforcement method in Section 4. In the second part of the paper, we develop and test the moment-based estimator for VARMA models in Section 5. Finally, we discuss the results of the numerical experiments and offer some conclusions in Section 6.

We will use the following notation. We use I_n (or just I) for the identity matrix of order n. We denote by $\overline{A}$, A^T, and A^* the conjugate, the transpose, and the conjugate transpose of a matrix A, respectively. The symbol $\rho(A)$ denotes the spectral radius of a matrix, i.e.,

$$\rho(A) = \max_{\lambda \text{ eigenvalue of } A} |\lambda|.$$

For a vector x we denote by $\|x\|$ its standard Euclidean norm. For a matrix A, $\|A\|_2 := (\rho(A^*A))^{1/2}$ denotes the spectral norm, whereas $\|A\|_F := (\sum_{i,j}|a_{ij}|^2)^{1/2}$ the Frobenius norm. The latter, used in error computations, relates directly to the notion of RMS (root mean square) error, common in the statistical community.

2 Spectral plots

Let $A, B \in \mathbb{R}^{n,n}$, $B = B^T$, and consider the matrix Laurent polynomial $Q(\lambda) = \lambda A + B + \lambda^{-1}A^T$. Such a polynomial is called *palindromic* because of the structure imposed on its coefficients: reversing the order of the coefficients results in the original Laurent polynomial, only transposed. We assume here and in the following that $Q(\lambda)$ is *regular*, that is, $\det Q(\lambda)$ does not vanish for each $\lambda \in \mathbb{C}$.

Observe that for all $\lambda \in \mathbb{C}$ with $\lambda \neq 0$ we have $\left(Q(\overline{\lambda}^{-1})\right)^* = Q(\lambda)$. Since for $\lambda = e^{\imath\omega}$ on the unit circle we have

$$\overline{\lambda}^{-1} = e^{-\overline{\imath}\overline{\omega}} = e^{\imath\omega} = \lambda,$$

we see that

$$(Q(e^{\imath\omega}))^* = Q(e^{\imath\omega}),$$

i.e., Q is Hermitian on the unit circle. We wish to perturb A and B such that the parameter-dependent matrix $\tilde{Q}(\lambda)$ (arising from $Q(\lambda)$ upon perturbation of A and B) is positive definite for each choice of the parameter λ on the unit circle.

A useful tool for visualizing the eigenvalues of Q on the unit circle is a spectral plot such as the one in Figure 1. On the x-axis we display the interval $[-\pi, \pi]$; for each value ω in

Figure 1: The spectral plot of a certain 4×4 matrix Laurent polynomial Q on the unit circle

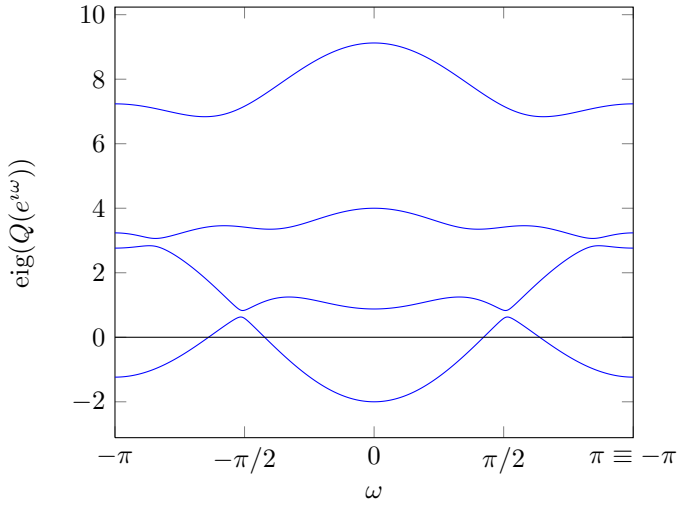
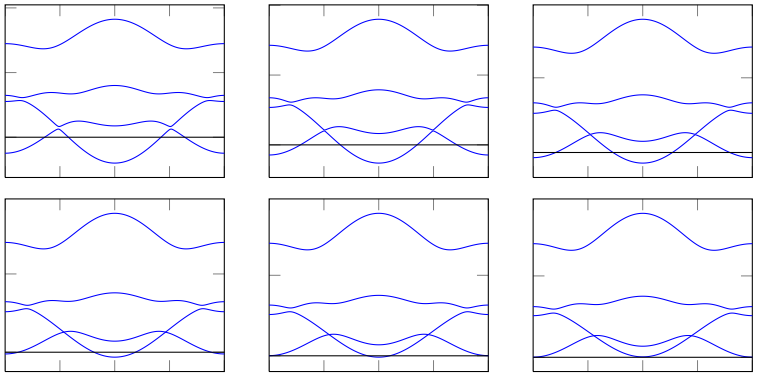


Figure 2: Successive iterations of the solvability enforcement procedure. In each iteration, we aim to construct a perturbation that shifts the generalized eigenvalues by a distance 0.2 in the direction indicated by their sign characteristic.



this interval, we plot the eigenvalues of the Hermitian matrix $Q(e^{\imath\omega})$. Obviously the plot is 2π-periodic and symmetric (because $Q(e^{\imath\omega}) = Q(e^{\imath(\omega+2\pi)}) = (Q(e^{-\imath\omega}))^T$). For instance, one sees from the graph that all the lines lie above the x-axis for $\omega = \pi/2$, so $Q(e^{\imath\pi/2})$ is positive definite. Instead, for $\omega = 0$ (and in fact for most values of ω) there is an eigenvalue below the imaginary axis, thus $Q(e^{\imath 0})$ is not positive definite. There are four points in which the lines cross the x-axis and these correspond to the values of ω for which $Q(e^{\imath\omega})$ is singular, i.e., for which $e^{\imath\omega}$ is a generalized eigenvalue of Q on the unit circle.

Notice that the lines corresponding to different eigenvalues in our spectral plot example come very close to each other, but never cross. This is not an error in the graph, but an instance of a peculiar phenomenon known as *eigenvalue avoidance*, see, e.g., [21, p. 140] or [3].

To understand what we need to do to enforce positivity, we make use of a heuristic argument based on this spectral plot. A continuous perturbation that renders $Q(\lambda)$ positive definite on the whole unit circle will have to move up the two bumps that extend below the x-axis. For instance, we display in Figure 2 six successive snapshots from a modification which has the effect of making the pencil positive definite without altering too much the shape of the eigenvalue curves.

Let us now consider what happens during this process to the four intersections between

the lines and the x-axis, which we now label ω_1, ω_2, ω_3, ω_4 starting from left. The two central intersections ω_2 and ω_3 move towards each other, until they coalesce and then disappear (i.e., the curve does not cross the x-axis anymore). The other two intersections ω_1 and ω_4 move towards the borders of the graph, coalesce at $\omega = \pi$ and then disappear as well. In particular, we see that the intersections ω_i in which the slope of the line crossing the x-axis is positive ($i = 1, 3$) need to be moved to the *left*, and the ω_i for which it is negative ($i = 2, 4$) need to be moved to the *right*. This is a crucial observation.

One usually defines the *(generalized) eigenvalues* and eigenvectors of the matrix polynomial Q as the pairs $(\lambda, v) \in \mathbb{C} \times \mathbb{C}^n$ such that $Q(\lambda)v = 0$ and $v \neq 0$. This is a different notion than that of the eigenvalues appearing in the spectral plot. To avoid confusion between the two, in this paper we do not drop the adjective *generalized*.

The spectral plot gives nevertheless some information on the generalized eigenvalues of Q: the matrix Laurent polynomial Q has a generalized eigenvalue $\lambda = e^{\imath\omega}$ on the unit circle if and only if $Q(e^{\imath\omega})$ is singular, i.e., a line in the spectral plot touches the real axis at $x = \omega$. From this observation, it is apparent that the presence of generalized eigenvalues of Q on the unit circle is no isolated phenomenon, but rather it is stable with respect to small structured perturbations, i.e., perturbations that keep the palindromic structure: if some lines cross the x-axis, small perturbations will not change this fact. Moreover, the sign of the slope with which the line crosses the x-axis is known in the literature as *sign characteristic* of the unimodular generalized eigenvalue [10], and it is well known that only two close-by generalized eigenvalues with opposite sign characteristics can move off the unit circle through small perturbations.

Our perturbation algorithm borrows from this intuition and consists of making several small perturbations to A and B incrementally, trying to move iteratively the unimodular generalized eigenvalues in the right direction according to their sign characteristic. Eventually, the generalized eigenvalues coalesce in pairs with opposite sign characteristics and then leave the unit circle.

In order to relate the change of the generalized eigenvalues to a certain small perturbation of the two matrices A, B, we rely on first-order perturbation theory for generalized eigenvalues of palindromic Laurent polynomials.

3 Palindromic perturbation theory

In this section, we survey some results on generalized eigenvalue perturbations which are useful for developing our regularization algorithm.

Theorem 1. *Let $A, B \in \mathbb{R}^{n,n}$ with $B = B^T$ and let $\lambda_0 \neq 0 \in \mathbb{C}_*$ be a simple generalized eigenvalue of the regular palindromic matrix Laurent polynomial $Q(\lambda) := \lambda A + B + \lambda^{-1}A^T$ with right eigenvector v_0 and left eigenvector u_0. Then, $u_0^* \left(A - \lambda_0^{-2}A^T\right) v_0 \neq 0$, and there exists a function $\Phi : \mathbb{R}^{n,n} \times \mathbb{R}^{n,n} \to \mathbb{C}$ which fulfills $\Phi(E, F) = o(\|E, F\|)$ for $(E, F) \to 0$ such that*

$$\tilde{\lambda}_0 := \lambda_0 - \frac{u_0^* \left(\lambda_0 E + F + \lambda_0^{-1}E^T\right) v_0}{u_0^* \left(A - \lambda_0^{-2}A^T\right) v_0} + \Phi(E, F) \tag{3}$$

is a generalized eigenvalue of the perturbed matrix Laurent polynomial

$$\lambda(A + E) + (B + F) + \lambda^{-1}(A + E)^T.$$

Proof. This is a special case of [27, Lemma 2.7], obtained by setting

$$T\left(\lambda, \begin{bmatrix} \mathrm{vec}\,(E) \\ \mathrm{vec}\,(F) \end{bmatrix}\right) = \lambda(A + E) + (B + F) + \lambda^{-1}(A + E)^T.$$

□

For generalized eigenvalues on the unit circle $\lambda_0 = e^{\imath\omega_0}$ we have that right eigenvectors v_0 (i.e., $Q(e^{\imath\omega_0})v_0 = 0$) also satisfy

$$0 = (Q(e^{\imath\omega_0})v_0)^* = v_0^* \left(Q(e^{\imath\omega_0})\right)^* = v_0^* Q(e^{\imath\omega}),$$

i.e., they are also left eigenvectors.

We wish to give an explicit expression for the slopes of the lines crossing the x-axis in the spectral plot. We first need to recall the following classical result in matrix perturbation theory [18, Section II.6.2]: given an analytical Hermitian matrix function $H(\omega)$ of a real parameter ω, one can choose an orthonormal basis of eigenvectors which is an analytical function of ω.

Lemma 2. *Let $A, B \in \mathbb{R}^{n,n}$ with $B = B^T$ and let $\lambda_0 = e^{\imath\omega_0}$ be a unimodular generalized eigenvalue of the matrix Laurent polynomial $Q(\lambda) = \lambda A + B + \lambda^{-1}A^T$. Let v_0 be a corresponding eigenvector; if λ_0 is a multiple eigenvalue, then v_0 needs to be chosen in such a way that it can be continued analytically as a function $v_0(\omega)$ in a neighborhood of ω_0.*

Then, the spectral plot for Q contains a line that crosses the x-axis at $(\omega_0, 0)$ with slope $\sigma_0 := -2\Im(\lambda_0 v_0^* A v_0)$.

Proof. As claimed above, the spectral plot passes through the point $(\omega_0, 0)$ since the matrix $Q(\lambda_0)$ is singular and thus it has a zero eigenvalue. As stated above, one can choose analytic functions $v(\omega)$ and $\theta(\omega)$ such that $v(\omega)$ is an eigenvector of $Q(e^{\imath\omega})$ with eigenvalue $\theta(\omega)$, $\|x(\omega)\| = 1$ for each ω, and $v(\omega_0) = v_0$, $\theta(\omega_0) = 0$.

The slope of the line through $(\omega_0, 0)$ is given by

$$\dot\theta(\omega_0) = \left.\frac{d\,\theta(\omega)}{d\,\omega}\right|_{\omega=\omega_0}.$$

We compute it by differentiating the relation $\theta(\omega) = v(\omega)^* Q(e^{\imath\omega}) v(\omega)$ with respect to ω and evaluating the resulting expression in ω_0. We get

$$\begin{aligned}\dot\theta(\omega_0) = \dot v(\omega_0)^* \underbrace{Q(e^{\imath\omega_0})v(\omega_0)}_{=\theta(\omega_0)v(\omega_0)=0} + \underbrace{v(\omega_0)^* Q(e^{\imath\omega_0})}_{=v(\omega_0)^*\theta(\omega_0)=0} \dot v(\omega_0) + v(\omega_0)^* \left.\frac{d\,Q(e^{\imath\omega})}{d\,\omega}\right|_{\omega=\omega_0} v(\omega_0)\\ = v_0^*(\imath e^{\imath\omega_0} A - \imath e^{-\imath\omega_0} A^T)v_0 = -2\Im(e^{\imath\omega_0} v_0^* A v_0). \quad \square\end{aligned}$$

Finally, we specialize Theorem 1 to unimodular generalized eigenvalues of a palindromic matrix Laurent polynomial.

Theorem 3. *Let $A, B \in \mathbb{R}^{n,n}$ with $B = B^T$ and let $Q(\lambda) = \lambda A + B + \lambda^{-1}A^T$ have a simple unimodular generalized eigenvalue $\lambda_0 = e^{\imath\omega_0}$, with eigenvector v_0 and spectral slope $\sigma_0 = -2\Im(\lambda_0 v_0^* A v_0)$. Let $\tilde Q(\lambda) := \lambda(A + E) + B + F + \lambda^{-1}(A + E)^T$ be a sufficiently small perturbation of $Q(\lambda)$, with $F = F^T$. Then $\sigma_0 \neq 0$ and $\tilde Q$ has a generalized eigenvalue $\tilde\lambda_0 = e^{\imath\tilde\omega_0}$ such that*

$$\sigma_0(\tilde\omega_0 - \omega_0) = -v_0^* F v_0 - 2\Re(e^{\imath\omega_0} v_0^* E v_0) + \hat\Phi(E, F). \tag{4}$$

for some function $\hat\Phi(E, F)$ with $\hat\Phi(E, F) = o(\|E, F\|)$.

Proof. First note that Theorem 1 implies that $\sigma_0 \neq 0$. Applying some algebraic manipulations to (3), we get

$$\begin{aligned} e^{\imath\tilde\omega_0} &= e^{\imath\omega_0} - \frac{v_0^*(e^{\imath\omega_0}E + F + e^{-\imath\omega_0}E^T)v_0}{v_0^*(A - e^{-2\imath\omega_0}A^T)v_0} + \Phi(E, F)\\ &= e^{\imath\omega_0} + e^{\imath\omega_0}\frac{v_0^* F v_0 + 2\Re(e^{\imath\omega_0} v_0^* E v_0)}{v_0^*(-e^{\imath\omega_0}A + e^{-\imath\omega_0}A^T)v_0} + \Phi(E, F)\\ &= e^{\imath\omega_0} + e^{\imath\omega_0}\frac{v_0^* F v_0 + 2\Re(e^{\imath\omega_0} v_0^* E v_0)}{\imath\sigma_0} + \Phi(E, F)\\ &= e^{\imath\omega_0}\left(1 + \frac{v_0^* F v_0 + 2\Re(e^{\imath\omega_0} v_0^* E v_0)}{\imath\sigma_0}\right) + \Phi(E, F).\end{aligned}$$

Taking complex logarithms and using the expansion $\log(1+\varepsilon) = \epsilon + o(\varepsilon)$ leads to

$$\imath\tilde{\omega}_0 = \imath\omega_0 + \frac{v_0^* F v_0 + 2\Re(e^{\imath\omega_0} v_0^* E v_0)}{\imath\sigma_0} + \tilde{\Phi}(E,F), \qquad \tilde{\Phi}(E,F) = o(\|E,F\|)$$

from which (4) follows. □

This establishes a first order relation between the change of a unimodular eigenvalue and the perturbation of the coefficients.

4 Solvability enforcement

In this section, we wish to derive a procedure for making a small-norm modification of a palindromic Laurent matrix polynomial that moves the unimodular generalized eigenvalues (approximately) in a different nearby location. The derivation is along the lines of [5] (with modifications to account for the different type of structure of the matrix polynomial).

Suppose that the matrix polynomial $Q(\lambda) = \lambda A + B + \lambda^{-1} A^T$ has unimodular generalized eigenvalues $\lambda_j = e^{\imath\omega_j}$, for $j = 1, 2, \ldots, \ell$. Let σ_j and v_j denote the spectral slope and the eigenvector associated to λ_j, normalized such that $\|v_j\| = 1$.

We are looking for a perturbed matrix polynomial $\tilde{Q}(\lambda) = \lambda(A+E) + B + F + \lambda^{-1}(A+E)^T$, with $\|E\|$ and $\|F\|$ small, that has generalized eigenvalues close to specified points $e^{\imath\tilde{\omega}_j}$, $j = 1, 2, \ldots, \ell$ on the unit circle. We assume that the $\tilde{\omega}_i$ are chosen in a way such that $|\tilde{\omega}_j - \omega_j|$ is small for all j.

Moreover, we wish to allow only perturbations in the special form

$$(E,F) = \sum_{i=1}^{m} (E_i, F_i)\delta_i \tag{5}$$

for some $\delta_i \in \mathbb{R}$, where $(E_i, F_i) \in \mathbb{R}^{n,n} \times \mathbb{R}^{n,n}$, with $F_i = F_i^T$ for each $i = 1, 2, \ldots, m$, is a suitably chosen basis of allowed modifications to the pair (A, B).

For instance, if $n = 2$, a natural choice for this perturbation basis is

$$\left(\left[\begin{smallmatrix}1&0\\0&0\end{smallmatrix}\right],0\right), \left(\left[\begin{smallmatrix}0&0\\0&1\end{smallmatrix}\right],0\right), \left(\left[\begin{smallmatrix}0&1\\0&0\end{smallmatrix}\right],0\right), \left(\left[\begin{smallmatrix}0&0\\1&0\end{smallmatrix}\right],0\right), \left(0,\left[\begin{smallmatrix}1&0\\0&0\end{smallmatrix}\right]\right), \left(0,\left[\begin{smallmatrix}0&0\\0&1\end{smallmatrix}\right]\right), \left(0,\left[\begin{smallmatrix}0&1\\1&0\end{smallmatrix}\right]\right). \tag{6}$$

This choice gives all possible perturbations on the entries of each matrix that preserve the symmetry of B. However, one can choose a different basis $(E_i, F_i)_{i=1}^m$ for the perturbation space, restricting the choice to perturbations having a stronger linear structure. We shall need this additional freedom in Section 5.3.

We define for each $m, n \in \mathbb{N}$ the operator $\operatorname{vec}(\cdot) : \mathbb{C}^{m,n} \to \mathbb{C}^{mn}$ that stacks the columns of the matrix in its argument, i.e.,

$$(\operatorname{vec}(M))_{(j-1)m+i} = M_{ij}.$$

It is well-known (e.g., [16]) that $\operatorname{vec}(AXB) = (B^T \otimes A)\operatorname{vec}(X)$ for each triple of matrices A, X, B of compatible size, where $\otimes$ denotes the Kronecker product.

Using these tools, we can rewrite (5) as

$$\begin{bmatrix}\operatorname{vec}(E)\\ \operatorname{vec}(F)\end{bmatrix} = \begin{bmatrix}\operatorname{vec}(E_1) & \cdots & \operatorname{vec}(E_m)\\ \operatorname{vec}(F_1) & \cdots & \operatorname{vec}(F_m)\end{bmatrix} \begin{bmatrix}\delta_1\\ \vdots\\ \delta_m\end{bmatrix}$$

and (4) as

$$\begin{aligned}\sigma_j(\tilde{\omega}_j - \omega_j) &\approx \begin{bmatrix}-2\Re(e^{\imath\omega_i}(v_j^T \otimes v_j^*)) & -(v_j^T \otimes v_j^*)\end{bmatrix} \begin{bmatrix}\operatorname{vec}(E)\\ \operatorname{vec}(F)\end{bmatrix}\\ &= \begin{bmatrix}-2\Re(e^{\imath\omega_i}(v_j^T \otimes v_j^*)) & -\Re(v_j^T \otimes v_j^*)\end{bmatrix} \begin{bmatrix}\operatorname{vec}(E)\\ \operatorname{vec}(F)\end{bmatrix}.\end{aligned}$$

The last equality holds because F is real symmetric, thus

$$(v_j^T \otimes v_j^*)\,\mathrm{vec}\,(F) = v_j^* F v_j = \Re(v_j^* F v_j) = \Re\left((v_j^T \otimes v_j^*)\,\mathrm{vec}\,(F)\right) = \Re\left((v_j^T \otimes v_j^*)\right)\mathrm{vec}\,(F)\,.$$

We set

$$\mathcal{A} = \begin{bmatrix} -\Re(e^{\imath\omega_1}(v_1^T \otimes v_1^*)) & -\Re(v_1^T \otimes v_1^*) \\ \vdots & \vdots \\ -\Re(e^{\imath\omega_\theta}(v_\theta^T \otimes v_\theta^*)) & -\Re(v_\theta^T \otimes v_\theta^*) \end{bmatrix} \begin{bmatrix} \mathrm{vec}\,(E_1) & \cdots & \mathrm{vec}\,(E_m) \\ \mathrm{vec}\,(F_1) & \cdots & \mathrm{vec}\,(F_m) \end{bmatrix},\ \mathcal{B} = \begin{bmatrix} \sigma_1(\tilde{\omega}_1 - \omega_1) \\ \vdots \\ \sigma_\ell(\tilde{\omega}_\ell - \omega_\ell) \end{bmatrix},$$

and $\delta = [\delta_1, \ldots, \delta_m]^T$, thus obtaining the system of ℓ linear equations in m unknowns

$$\mathcal{A}\delta = \mathcal{B}. \tag{7}$$

So, any sufficiently small perturbation (5) satisfying (7) moves the unimodular generalized eigenvalues approximately to the wanted positions. We are interested in the smallest such perturbation. To this end we assume that system (7) is underdetermined, $m > \ell$, but full-rank. Hence, we can use a simple QR factorization to compute its minimum-norm solution. In fact (e.g., [11]), the minimum norm solution is given by $\delta = QR^{-T}\mathcal{B}$, where $\mathcal{A}^T = QR$ denotes a thin QR factorization.

To sum up, our regularization algorithm is as follows.

Algorithm 1.
Input: $A, B = B^T \in \mathbb{R}^{n,n}$ such that $Q(\lambda) = \lambda A + B + \lambda^{-1}A^T$ is regular.
Output: $\tilde{A}, \tilde{B} = \tilde{B}^T \in \mathbb{R}^{n,n}$ such that $\tilde{Q}(\lambda) = \lambda\tilde{A} + \tilde{B} + \lambda^{-1}\tilde{A}^T$ has no unimodular eigenvalues.

1. $\tilde{A} \leftarrow A, \tilde{B} \leftarrow B$
2. Compute the unimodular generalized eigenvalues $\lambda_j = e^{\imath\omega_j}$ of $\tilde{Q}(\lambda)$, $j = 1, 2, \ldots, \ell$ and the associated eigenvectors v_j. If there is none, terminate the algorithm. Also compute the spectral slopes $\sigma_j = 2\Im(\lambda_j v_j^* \tilde{A} v_j^*)$.
3. Determine suitable locations for the perturbed generalized eigenvalues $\tilde{\omega}_j$ "in the right direction" and not too far away from the ω_j. For instance, we used in the following experiments $\tilde{\omega}_j = \omega_j - \tau\,\mathrm{sign}(\sigma_j)$, with step size $\tau = 10^{-2}$ (unless specified otherwise). Other choices are discussed in [5].
4. Assemble the system (7) and compute its minimum-norm solution δ.
5. Set $\tilde{A} \leftarrow \tilde{A} + \sum_{i=1}^m \delta_i E_i$, $\tilde{B} \leftarrow \tilde{B} + \sum_{i=1}^m \delta_i F_i$ and repeat from step 2.

Typically, several iterations of this procedure are needed to achieve a matrix polynomial with no unimodular generalized eigenvalues. The smaller the step size τ is, the more iterations are needed, but the smaller is the resulting total perturbation to A and B. In any case, all these results come from first-order approximations, so there is no formal guarantee that what we are computing is the *smallest* possible perturbation that makes $Q(\lambda)$ positive definite on the unit circle.

4.1 A numerical example

As a first example of our solvability enforcement procedure, we describe in more detail the experiment already portrayed in Figures 1 and 2. We start from the matrices

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}, \quad B = \begin{bmatrix} 3 & 2 & 1 & 0 \\ 2 & 3 & 2 & 1 \\ 1 & 2 & 3 & 2 \\ 0 & 1 & 2 & 3 \end{bmatrix}, \tag{8}$$

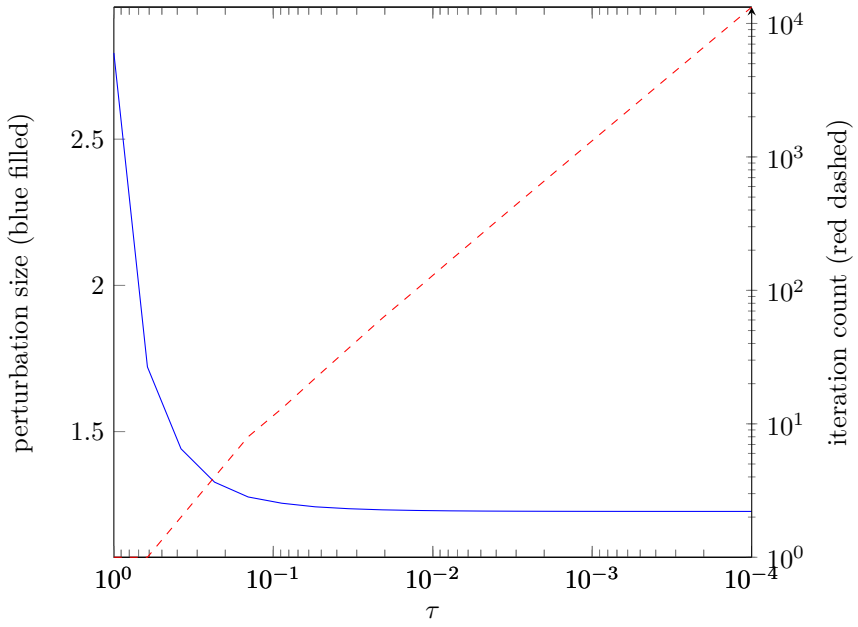


Figure 3: Absolute magnitude of the final perturbation, $\|[\tilde{A} - A, \tilde{B} - B]\|_F$, and number of iterations in the solvability enforcement procedure for different values of the step size τ

for which the resulting spectral plot was given in Figure 1. We ran Algorithm 1 with a step size of $\tau = 0.2$; the step-size was increased in this experiment to reduce the number of needed steps and make Figure 2 clearer. We reported the spectral plot for the resulting polynomial $\tilde{Q}(\lambda)$ after each iteration of the procedure in Figure 2. One can see that the profile of the topmost eigenvalue curves is almost unchanged, while the bottom ones are modified slightly at each iteration, and these modifications have the overall effect of slowly pushing them upwards. After six iterations, the resulting palindromic Laurent matrix polynomial $Q(\lambda)$ is positive definite on the whole unit circle (as shown in the bottom-right graph), and thus it has no more unimodular generalized eigenvalues. The resulting matrices returned by the algorithm are

$$\tilde{A} \approx \begin{bmatrix} 0.816 & 0.183 & 0.0379 & -0.0565 \\ 0.183 & 0.915 & 0.775 & 0.152 \\ 0.0379 & 0.775 & -0.647 & -0.173 \\ -0.0565 & 0.152 & -0.173 & -0.922 \end{bmatrix}, \quad \tilde{B} \approx \begin{bmatrix} 3.16 & 1.67 & 0.956 & 0.0913 \\ 1.67 & 3.28 & 1.62 & 1.13 \\ 0.956 & 1.62 & 3.41 & 1.55 \\ 0.0913 & 1.13 & 1.55 & 3.13 \end{bmatrix}.$$

The relative magnitude of the obtained perturbation is

$$\frac{\|\tilde{A} - A\|_F}{\|A\|_F} = 0.2755, \quad \frac{\|\tilde{B} - B\|_F}{\|B\|_F} = 0.1398.$$

Such a large value is not unexpected, since the plot in Figure 1 extends significantly below the real axis, thus quite a large perturbation is needed.

The step size τ plays a role here. With a smaller value of τ, one expects the resulting perturbation to be smaller, since the approximated first-order eigenvalue locations are interpolated more finely; on the other hand, the number of needed steps should increase as well. This expectation is supported by Figure 3 where we report the resulting values for different choices of τ in this example.

More experiments from applications are discussed in the following sections.

5 Parameter estimation for Weak VARMA(1,1) models

Here we show how solvability enforcement of the matrix equation (1) can improve parameter estimation for VARMA(1,1) models. A multivariate time series (i.e., a sequence of d-variate probability variables) $(x_t)_{t=1,2,\ldots}$ in $\mathbb{R}^d$ obeys a *weak VARMA(1,1) model* (vector autoregressive model with moving average) if it satisfies a relation of the form

$$x_t - \Phi x_{t-1} = c + u_t - \Theta u_{t-1}, \quad t = 2, 3, \ldots \tag{9}$$

with parameters $\Phi, \Theta \in \mathbb{R}^{d,d}$, and $c \in \mathbb{R}^d$. Here, (u_t) is a d-variate weak white noise process, i.e., it satisfies

1. $\mathbb{E}[u_t] = 0$ for all t (zero mean);
2. $\mathbb{E}[u_t u_{t+j}^T] = 0$ for all t and all $j \neq 0$ (uncorrelatedness).

Here and in the following, $\mathbb{E}[\cdot]$ denotes the expected value of a random variable. Moreover, we adopt the convention that the observed or estimated value of a random variable or exact asymptotic quantity is denoted by the same letter with a hat; so, for instance, x_t is a multivariate random variable while $\hat{x}_t$ is a real vector that corresponds to a realization of x_t.

The task of estimating the parameters Φ, Θ, and c of a VARMA(1,1) model, given a finite number N of observations $\hat{x}_1, \ldots, \hat{x}_N \in \mathbb{R}^d$, is more challenging than the corresponding problem for the more common model with $\Theta = 0$ (known as *AR(1)* model).

The dominant approach is *maximum likelihood* (ML) estimation which can be applied when $u_1, u_2, \ldots$ are independent and identically distributed random variables. The ML estimator is defined as

$$(\hat{\Phi}_{ML}, \hat{\Theta}_{ML}) := \arg\max L(\hat{\Phi}, \hat{\Theta}), \tag{10}$$

where $L(\hat{\Phi}, \hat{\Theta})$ denotes the *likelihood* of the estimates $\hat{\Phi}$, $\hat{\Theta}$ (and of $\hat{x}_1$ if it is not provided by the application), that is, the probability density that a process satisfying (9) (with Φ, Θ replaced by $\hat{\Phi}, \hat{\Theta}$) generated the observed time series $(\hat{x}_t)_{t=1}^N$. The main issue with maximum likelihood estimation is that the optimization problem (10) is very complex to solve numerically. It is generally a non-convex problem, and the number of parameters grows quadratically with the dimension d of the time series. Standard optimization methods may require thousands of iterations to converge [17], and each of these iterations requires at least one pass through all the time series data $\hat{x}_1, \ldots, \hat{x}_N$. Moreover, most ML estimation approaches assume Gaussian independent noise u_t, tempting practitioners to run them blindly even when the noise process is known *not* to be Gaussian independent. Admittedly, doing so is not completely unjustified, as in this case the real likelihood function can be replaced by the one that would be obtained if the u_t were Gaussian independent, leading to the *quasi-maximum likelihood* approach [22, Section 16.4]. However, the obtained asymptotic results are considerably weaker, since replacing the likelihood function is only a heuristic procedure.

We suggest here instead a *moment-based estimator* (also known as *method of moments*), that is, an estimator that is based on computing the moments of the time series and solving specific equations involving them. Moment-based estimators are known for a wide variety of statistical models [22]. Their precision is typically inferior to that of ML estimators, but they are simpler and they require less assumptions on the underlying probability distribution. Moreover, they can be used to provide good initial estimates for maximum likelihood routines. Moment-based estimators have been proposed in the literature for univariate ARMA models ($d = 1$) [20, 24], in which they require simply solving a (scalar) quadratic equation. An extension to multivariate models, which requires more advanced linear algebra techniques, was suggested in [26] for a specific econometric model, the multivariate GARCH(1,1) model; we describe it here for the more general case of a weak VARMA(1,1) model. Our moment-based estimator requires only the mean of the time series and the first three autocorrelations, which can all be approximated with just one pass through the data. It is therefore orders of magnitude faster than an ML estimator, and thus feasible even for large data sets.

Additional assumptions We make the following classical additional assumptions on the weak VARMA(1,1) process. We recall that we denote the spectral radius of a matrix M by $\rho(M)$.

1. $\rho(\Phi) < 1$ (*asymptotic stationarity*) and $\rho(\Theta) < 1$ (*invertibility*).
2. The process is *ergodic*, i.e., for any measurable function f, the average over time $\frac{1}{N}\sum_{t=1}^{N} f(\hat{x}_t)$ converges for almost all possible observations $\hat{x}_t$ of the random variables x_t to a value (possibly ∞) that is also the limit of $\mathbb{E}[f(x_t)]$ for $t \to \infty$. We adopt the shorthand $\mathbb{E}_t[x_t] := \lim_{t\to\infty} \mathbb{E}[x_t]$ to denote the stationary limit.
3. The variance of the noise $\mathbb{E}[u_t u_t^T]$ is finite.

In particular, we do not assume that the u_t are independent (but only that they are uncorrelated, see above), nor that they have the same law, nor a specific distribution such as the Gaussian one.

5.1 Estimating Φ and c

We define the asymptotic mean $\mu := \mathbb{E}_t[x_t]$ and for $k = 0, 1, \ldots$ the asymptotic autocovariances $M_k := \mathbb{E}_t[\bar{x}_{t+k}\bar{x}_t^T]$, where $\bar{x}_t := x_t - \mu$. It follows from our assumptions that μ and M_k, $k = 0, 1, 2, \ldots$, exist and are finite.

We have by (9)

$$(I - \Phi)\mu = \mathbb{E}_t[x_t - \Phi x_{t-1}] = \mathbb{E}_t[c + u_t - \Theta u_{t-1}] = c.$$

This allows us to write a version of (9) without the term c,

$$\bar{x}_t - \Phi\bar{x}_{t-1} = u_t - \Theta u_{t-1}.$$

Hence, for $k \geq 2$ we have the Yule-Walker relation

$$M_k = \mathbb{E}_t[\bar{x}_{t+k}\bar{x}_t^T] = \mathbb{E}_t[(\Phi\bar{x}_{t+k-1} + u_{t+k} - \Theta u_{t+k-1})\bar{x}_t^T] = \mathbb{E}_t[\Phi\bar{x}_{t+k-1}\bar{x}_t^T] = \Phi M_{k-1}.$$

Here we used that for these k the terms $\mathbb{E}_t[u_{t+k}\bar{x}_t^T]$ and $\mathbb{E}_t[u_{t+k-1}\bar{x}_t^T]$ vanish, because u_{t+k-1} and u_{t+k} are uncorrelated with the values of u_j for all $j \leq t$ (and thus also with $\bar{x}_j$).

This property allows estimating Φ and c easily: since asymptotically consistent estimates of μ and M_k are readily available from the data in the form of $\hat{\mu}$ and $\hat{M}_k$ by (2) we can compute for instance $\hat{\Phi} = \hat{M}_2\hat{M}_1^{-1}$ and $\hat{c} = (I - \hat{\Phi})\hat{\mu}$.

Remark 4. An alternative strategy, using all the available finite-sample moments, is choosing weights $(w_i)_{i=2}^N \geq 0$ and solving in the least squares sense the overdetermined system

$$\begin{bmatrix} w_2\hat{M}_2 & w_3\hat{M}_3 & \cdots & w_N\hat{M}_N \end{bmatrix} = \hat{\Phi}\begin{bmatrix} w_2\hat{M}_1 & w_3\hat{M}_2 & \cdots & w_N\hat{M}_{N-1} \end{bmatrix}.$$

However, it is not clear what the best choice for the weights w_i is. As a consequence, the accuracy of the different estimates obtained by $\hat{M}_k\hat{M}_{k-1}^{-1}$ may vary wildly, and the experiments do not show a visible pattern for their distribution. We do not pursue this approach further.

5.2 Estimating Θ

Estimating the remaining parameter Θ requires more effort. The following strategy was suggested in [26] for another econometric model, but it can be applied to a generic weak VARMA(1,1) model with minor modifications.

Let $\Sigma := \mathbb{E}_t[u_t u_t^T]$ be the asymptotic covariance matrix of u_t and define $z_t := \bar{x}_t - \Phi\bar{x}_{t-1} = u_t - \Theta u_{t-1}$. Then the asymptotic covariance matrix Γ_0 of z_t can be computed in two different ways as

$$\begin{aligned}\Gamma_0 := \mathbb{E}_t[z_t z_t^T] &= \mathbb{E}_t[(\bar{x}_t - \Phi\bar{x}_{t-1})(\bar{x}_t - \Phi\bar{x}_{t-1})^T] = M_0 - \Phi M_1^T - M_1\Phi^T + \Phi M_0\Phi^T, \\ &= \mathbb{E}_t[(u_t - \Theta u_{t-1})(u_t - \Theta u_{t-1})^T] = \Sigma + \Theta\Sigma\Theta^T.\end{aligned}$$

In the second expression we used the fact that $\mathbb{E}[u_t u_{t-1}^T] = 0$, since noises at different times are uncorrelated. Similarly, for $\Gamma_1 := \mathbb{E}_t[z_{t+1} z_t^T]$ we obtain $\Gamma_1 = M_1 - \Phi M_0 = -\Theta\Sigma$.

The first expressions for Γ_0 and Γ_1 can be used to compute their estimates $\hat{\Gamma}_0$ and $\hat{\Gamma}_1$ from $\hat{M}_k$ and $\hat{\Phi}$ as

$$\begin{aligned} \hat{\Gamma}_0 &= \hat{M}_0 - \hat{\Phi}\hat{M}_1^T - \hat{M}_1\hat{\Phi}^T + \hat{\Phi}\hat{M}_0\hat{\Phi}^T, \\ \hat{\Gamma}_1 &= \hat{M}_1 - \hat{\Phi}\hat{M}_0. \end{aligned} \tag{11}$$

The second expressions for Γ_0 and Γ_1,

$$\begin{aligned} \Gamma_0 &= \Sigma + \Theta\Sigma\Theta^T, \\ \Gamma_1 &= -\Theta\Sigma, \end{aligned}$$

may, using simple algebraic manipulations, be transformed into two decoupled equations in Σ and $Y := \Theta^T$, respectively:

$$\Gamma_0 = \Sigma + \Gamma_1\Sigma^{-1}\Gamma_1^T, \tag{12}$$

$$\Gamma_1 Y^2 + \Gamma_0 Y + \Gamma_1^T = 0. \tag{13}$$

The last equation (13) is a quadratic matrix equation. The following result shows how a solution $Y = \Theta^T$ can be constructed using generalized eigenpairs.

Theorem 5. *[9, Section 4.2] Let $\Gamma_0, \Gamma_1 \in \mathbb{R}^{d,d}$ with $\Gamma_0 = \Gamma_0^T$ be such that the matrix Laurent polynomial $\Gamma(\lambda) := \lambda\Gamma_1 + \Gamma_0 + \lambda^{-1}\Gamma_1^T$ is regular, and has $2d$ distinct generalized eigenvalues. Then, each solution of the matrix equation (13) can be written in the form*

$$Y = \begin{bmatrix} v_1 & v_2 \dots & v_d \end{bmatrix} \operatorname{diag}(\lambda_1, \lambda_2, \dots, \lambda_d) \begin{bmatrix} v_1 & v_2 \dots & v_d \end{bmatrix}^{-1}, \tag{14}$$

where (λ_j, v_j), $j = 1, 2, \dots, d$, are d of the $2d$ generalized eigenpairs of Γ.

The generalized eigenpairs of Γ (necessary for the construction of Y) can be approximated by generalized eigenpairs of $\hat{\Gamma}(\lambda) := \lambda\hat{\Gamma}_1 + \hat{\Gamma}_0 + \lambda^{-1}\hat{\Gamma}_1$. These in turn can be computed by several approaches, e.g., using a Schur form (or structured Schur form [19, 28]) of a linearization of the polynomial $\lambda\Gamma(\lambda)$ [15, 23] or by doubling-type algorithms [25].

Summarizing the results so far, a first preliminary VARMA estimator could consist of computing i) $\hat{\mu}$, $\hat{M}_0$, $\hat{M}_1$, $\hat{M}_2$ by (2); ii) $\hat{\Phi} = \hat{M}_2\hat{M}_1^{-1}$ and $\hat{c} = (I - \hat{\Phi})\hat{\mu}$; iii) $\hat{\Gamma}_0, \hat{\Gamma}_1$ by (11); iv) d eigenpairs of $\hat{\Gamma}$ by any suitable method; and v) $\hat{\Theta}$ as Y^T in (14).

An issue that remains is whether the such obtained Θ satisfies $\rho(\Theta) < 1$ (as required by our additional assumption 1). Another issue is the solvability of (12). By our derivation (12) is solvable, because Σ is known to be a solution. However, we replaced Γ_0 and Γ_1 by their estimates $\hat{\Gamma}_0$ and $\hat{\Gamma}_1$, which results in a perturbation of the intended equation. Then solvability has to be assessed by other means. We will use the following theorem, which is a minor modification of the results in [8].

Theorem 6. *Let $\Gamma_0, \Gamma_1 \in \mathbb{R}^{d,d}$ with $\Gamma_0 = \Gamma_0^T$. Then there exist matrices $\Sigma \in \mathbb{R}^{d\times d}$ with $\Sigma > 0$ and $Y = -\Sigma^{-1}\Gamma_1^T$ with $\rho(Y) < 1$ that solve (12) and (13), respectively, if and only if the parametric matrix $\Gamma(\lambda) := \lambda\Gamma_1 + \Gamma_0 + \lambda^{-1}\Gamma_1^T$ is Hermitian positive definite for every λ on the unit circle.*

In this case, Γ has exactly d generalized eigenvalues (counted with algebraic multiplicity) inside the unit circle.

In light of Theorem 6 all is well whenever $\hat{\Gamma}_0$ and $\hat{\Gamma}_1$ is such that $\hat{\Gamma}$ does not have unimodular generalized eigenvalues. Unfortunately, this cannot be guaranteed; unimodular generalized eigenvalues may exist. In this case the estimates $\hat{\Gamma}_0$ and $\hat{\Gamma}_1$ cannot be considered accurate and thus it is justified to perturb them slightly. The method described in Section 4 can be used to obtain nearby $\tilde{\Gamma}_0$ and $\tilde{\Gamma}_1$ such that $\tilde{\Gamma}(\lambda) := \lambda\tilde{\Gamma}_1 + \tilde{\Gamma}_0 + \lambda^{-1}\tilde{\Gamma}_1$ is without unimodular generalized eigenvalues. This amounts to inserting an extra step in above preliminary algorihm inbetween steps iii) and iv) consisting of computing $\tilde{\Gamma}_0$ and $\tilde{\Gamma}_1$ from $\hat{\Gamma}_0$ and $\hat{\Gamma}_1$ using Algorithm 1.

We will improve this estimator further in the following subsection.

5.3 Inflation and final algorithm

The perturbation technique described in Section 4 depends on a choice of linear perturbations (E_i, F_i), $i = 1, 2, \ldots, m$. It is natural to use componentwise unit perturbations in each entry of $\hat{\Gamma}_0$ and $\hat{\Gamma}_1$, i.e., perturbations with one or two 1 entries and all the other entries zero, as in (6). In this way, if we apply the result to the $d \times d$ Laurent polynomial $\lambda\hat{\Gamma}_1 + \hat{\Gamma}_0 + \lambda^{-1}\hat{\Gamma}_1^T$, we obtain perturbed matrices $\tilde{\Gamma}_0$, $\tilde{\Gamma}_1$ which are close to $\hat{\Gamma}_0$ and $\hat{\Gamma}_1$ in the Frobenius norm.

However, Γ_0 and Γ_1 are rather artificial functions of the moments, and it is not clear why one should be concerned in keeping the perturbation on them small. It seems a more sound strategy to search matrices $\tilde{M}_i$ that are close in norm to the estimated moments of the time series $\hat{M}_i$. This is not easy to do within our approach, since the $\hat{\Gamma}_i$ are nonlinear functions of the $\hat{M}_i$, and thus, even using first-order approximations only, the linearized perturbation basis (E_i, F_i) would be a rather complicated function of the $\hat{M}_i$.

We suggest an alternative strategy instead that computes Y from the generalized eigenvalues of an inflated matrix Laurent polynomial of dimension $3d \times 3d$. A similar construction was used in [6]. We state the following theorem in a slightly more general form than required, by including a free parameter k. In this paper we will use $k = 1$ only.

Theorem 7. *Let $k \in \mathbb{N}$, $k \geq 1$. Let $\hat{M}_0, \hat{M}_1, \hat{M}_k, \hat{M}_{k+1} \in \mathbb{R}^{d\times d}$ with $\hat{M}_0 = \hat{M}_0^T$ be given matrices, with $\hat{M}_k$ invertible. Set $\hat{\Phi} := \hat{M}_{k+1}\hat{M}_k^{-1}$ and define $\hat{\Gamma}_0$ and $\hat{\Gamma}_1$ as in* (11)*. Let (λ_i, v_i), $i = 1, 2, \ldots, 2d$, be the generalized eigenpairs of the matrix Laurent polynomial $\hat{\Gamma}(\lambda) := \lambda\hat{\Gamma}_1 + \hat{\Gamma}_0 + \lambda^{-1}\hat{\Gamma}_1^T$.*

Then, the matrix Laurent polynomial

$$\hat{R}(\lambda) := \lambda \begin{bmatrix} \hat{M}_1 & 0 & 0 \\ \hat{M}_0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} \hat{M}_0 & \hat{M}_1 & \hat{M}_{k+1} \\ \hat{M}_1^T & \hat{M}_0 & \hat{M}_k \\ \hat{M}_{k+1}^T & \hat{M}_k^T & 0 \end{bmatrix} + \lambda^{-1} \begin{bmatrix} \hat{M}_1^T & \hat{M}_0^T & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

has:

- *$2d$ generalized eigenvalues λ_i, $i = 1, \ldots, 2d$, where the λ_i are as defined above, with eigenvectors*

$$\begin{bmatrix} v_i \\ * \\ * \end{bmatrix}$$

  respectively, where by $*$ we denote a $d \times 1$ block whose content is not of interest here;

- *$2d$ generalized eigenvalues $\lambda_i = 0$, $i = 2d + 1, \ldots, 4d$, with eigenvectors given by the columns of*

$$\begin{bmatrix} 0 & 0 \\ I_d & 0 \\ 0 & I_d \end{bmatrix}; \tag{15}$$

- *$2d$ generalized eigenvalues $\lambda_i = \infty$, $i = 4d + 1, \ldots, 6d$, with eigenvalues given once again by the columns of* (15)*.*

We recall that the eigenvectors of a quadratic matrix polynomial need not be linearly independent, or even distinct, so it should not be surprising that the eigenvalues 0 and ∞ share the same eigenvectors.

Proof. Note that

$$\begin{bmatrix} \hat{M}_0 & \hat{M}_k \\ \hat{M}_k^T & 0 \end{bmatrix}^{-1} = \begin{bmatrix} 0 & \hat{M}_k^{-T} \\ \hat{M}_k^{-1} & -\hat{M}_k^{-1}\hat{M}_0\hat{M}_k^{-T} \end{bmatrix},$$

and (substituting $\hat{\Phi}$ by $\hat{M}_{k+1}\hat{M}_k^{-1}$ in (11))

$$\begin{aligned} \hat{\Gamma}_0 &= \hat{M}_0 - \hat{M}_1\hat{M}_k^{-T}\hat{M}_{k+1}^T - \hat{M}_{k+1}\hat{M}_k^{-1}\hat{M}_1^T + \hat{M}_{k+1}\hat{M}_k^{-1}\hat{M}_0\hat{M}_k^{-T}\hat{M}_{k+1}^T, \\ \hat{\Gamma}_1 &= \hat{M}_1 - \hat{M}_{k+1}\hat{M}_k^{-1}\hat{M}_0. \end{aligned}$$

Define

$$\begin{aligned} L(\lambda) &:= \begin{bmatrix} I_d & 0 \\ -\begin{bmatrix} \hat{M}_0 & \hat{M}_k \\ \hat{M}_k^T & 0 \end{bmatrix}^{-1} \begin{bmatrix} \hat{M}_1^T + \lambda \hat{M}_0 \\ \hat{M}_{k+1}^T \end{bmatrix} & I_{2d} \end{bmatrix} \\ &= \begin{bmatrix} I_d & 0 \\ -\begin{bmatrix} 0 & \hat{M}_k^{-T} \\ \hat{M}_k^{-1} & -\hat{M}_k^{-1}\hat{M}_0\hat{M}_k^{-T} \end{bmatrix} \begin{bmatrix} \hat{M}_1^T + \lambda \hat{M}_0 \\ \hat{M}_{k+1}^T \end{bmatrix} & I_{2d} \end{bmatrix} \\ &= \begin{bmatrix} I_d & 0 & 0 \\ -\hat{\Phi}^T & I_d & 0 \\ -(\hat{M}_k^{-1}(\hat{M}_1^T + \lambda \hat{M}_0) - \hat{M}_k^{-1}\hat{M}_0\Phi^T) & 0 & I_d \end{bmatrix}. \end{aligned}$$

One can check explicitly that

$$(L(\tfrac{1}{\lambda}))^T \hat{R}(\lambda) L(\lambda) = \begin{bmatrix} \hat{\Gamma}(\lambda) & 0 & 0 \\ 0 & \hat{M}_0 & \hat{M}_k \\ 0 & \hat{M}_k^T & 0 \end{bmatrix}.$$

As the latter matrix Laurent polynomial is block diagonal, it is easy to see that it has $2d$ generalized eigenvalues at zero and $2d$ at ∞, both with eigenvectors given by the columns of (15), and $2d$ generalized eigenvalues which coincide with the generalized eigenvalues λ_i of $\hat{Q}(\lambda)$, with eigenvectors

$$\begin{bmatrix} v_i \\ 0 \\ 0 \end{bmatrix}$$

respectively. Multiplying them by $L(\lambda)^{-1}$, we get the generalized eigenvalues of $\hat{R}(\lambda)$. In particular, the multiplication preserves the top block containing v_i. Also (15) is preserved. □

Theorem 7 shows that we can apply the positivity enforcement technique to $\hat{R}(\lambda)$ instead of to $\hat{\Gamma}(\lambda)$; the additional 0 and ∞ generalized eigenvalues are far away from the unit circle and pose no problem. Since the coefficients of $\hat{R}(\lambda)$ (in contrast to those of $\hat{\Gamma}(\lambda)$) are linear functions of the matrices $\hat{M}_i$, it is easy to construct a perturbation basis that corresponds to perturbing each entry of each $\hat{M}_i$ separately. One could say that our application is one of the rare cases in which a 200% inflation is beneficial in economics.

We have now derived our complete moment-based estimator which is stated below.

Algorithm 2.
Input: $\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_N \in \mathbb{R}^d$
Output: Estimates $\hat{c}, \hat{\Phi}, \hat{\Theta}$ for the parameters c, Φ, Θ of a weak VARMA(1,1) model that could have generated $\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_N$

1. Compute the finite-sample moments $\hat{\mu}$, $\hat{M}_0$, $\hat{M}_1$, $\hat{M}_2$ using (2).
2. Set
$$A \leftarrow \begin{bmatrix} \hat{M}_1 & 0 & 0 \\ \hat{M}_0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad B \leftarrow \begin{bmatrix} \hat{M}_0 & \hat{M}_1 & \hat{M}_2 \\ \hat{M}_1^T & \hat{M}_0 & \hat{M}_1 \\ \hat{M}_2^T & \hat{M}_1^T & 0 \end{bmatrix}.$$
3. Compute a perturbation basis (E_i, F_i) for (A, B) that corresponds to perturbing each entry of each $\hat{M}_k$, $k = 0, 1, 2$, separately.
4. Apply the perturbation technique described in Section 4, Algorithm 1 with the perturbation basis as chosen above, to obtain a nearby $\tilde{A}, \tilde{B}$ associated to $\tilde{M}_i$, $i = 0, 1, 2$, with no generalized eigenvalues on the unit circle.

5. Compute the generalized eigenpairs of $\lambda\tilde{A}+\tilde{B}+\lambda^{-1}\tilde{A}^T$ inside the unit circle; ignore those with eigenvectors in the form (15), and call the d remaining ones $(\lambda_j, w_j), j = 1, \ldots, d$.
6. Let $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_d)$ and V the upper $d \times d$ block of the matrix $\begin{bmatrix} w_1 & w_2 & \cdots & w_d \end{bmatrix}$.
7. Compute $\hat{\Theta} = Y^T = (VDV^{-1})^T$.
8. Compute $\hat{\Phi} = \tilde{M}_2\tilde{M}_1^{-1}$ and $\hat{c} = (I - \hat{\Phi})\hat{\mu}$.

Remark 8. *Consistency* is an important statistical property of an estimator producing an estimate $\hat{M}$ of a quantity M based on N observations. In the statistical literature, the estimator (and informally also the estimate $\hat{M}$ itself) is said to be *consistent* if $\lim_{N\to\infty} \hat{M} = M$ for almost all possible observations. The finite-sample moments $\hat{\mu}$ and $\hat{M}_k$ are consistent by ergodicity, and the moment-based estimator constructed here (without solvability enforcement) is an analytic function of these moments (provided $\rho(B) < 1$), hence it is consistent as well. Moreover, consistency (being an asymptotic property) is not affected by solvability enforcement at all, since solvability enforcement is not needed if the finite-sample moments are sufficiently close to their asymptotic values, which is the case for sufficiently large N. Thus our moment-based estimator, Algorithm 2, is consistent. As mentioned above, it is also computationally inexpensive. As a consequence, the moment-based estimates make an attractive choice as initial values for iterative implementations of more complicated estimators (such as the above-mentioned quasi-maximum likelihood estimator).

5.4 Numerical experiments

We considered two test cases for our experiments; both are VARMA(1,1) models (9) with Gaussian independent noise u_t with mean 0 and variance I. For this kind of noise and for given parameters c, Φ, Θ we can produce a time series by a simple MATLAB program like

```
function [X,U]=simulateVARMA(c,Phi,Theta,N,d)
X(:,1)=c; U(:,1)=randn(d,1);
for t=2:N
  U(:,t)=randn(d,1);
  X(:,t)=Phi*X(:,t-1)+c+U(:,t)-Theta*U(:,t-1);
end
```

where **randn** is MATLAB's buit-in normal distributed random number generator. We then construct the finite sample moments $\hat{\mu}$, $\hat{M}_0$, $\hat{M}_1$, $\hat{M}_2$ by (2) and the parameter estimates $\hat{c}, \hat{\Phi}, \hat{\Theta}$ by Algorithm 2 and assess the quality of these estimates as a function of the number of observations N. In particular we consider the following quantities:

- the relative errors of the finite-sample moments to the exact moments,

$$\mu_{\mathrm{err}} := \frac{\|\mu - \hat{\mu}\|}{\|\mu\|}, \quad M_{k,\mathrm{err}} := \frac{\|M_k - \hat{M}_k\|_F}{\|M\|_F}.$$

  For moment-based methods the quality of the finite-sample moments is crucial. We cannot expect high quality parameter estimates form low quality moment estimates.

- the errors of the parameter estimates, given by

$$c_{\mathrm{err}} := \frac{\|\hat{c} - c\|}{\mathbb{E}_t[\|x_t\|]}, \quad \Phi_{\mathrm{err}} := \|\hat{\Phi} - \Phi\|_2, \quad \text{and} \quad \Theta_{\mathrm{err}} := \|\hat{\Theta} - \Theta\|_2 \frac{\mathbb{E}_t[\|u_t\|]}{\mathbb{E}_t[\|x_t\|]}.$$

  The motivation to measure these errors like this stems from an important application. Typically, one is interested in using the obtained parameter estimates to predict a value for x_t as

$$x_t^P := \hat{c} + \hat{\Phi}x_{t-1} + u_t - \hat{\Theta}u_{t-1}.$$

Subtracting (9), one can bound the absolute prediction error as

$$\|x_t^P - x_t\| \leq \|\hat{c} - c\| + \|\hat{\Phi} - \Phi\|_2 \|x_{t-1}\| + \|\hat{\Theta} - \Theta\|_2 \|u_{t-1}\|.$$

Taking expectations and scaling by $\mathbb{E}_t[\|x_t\|]$ we have

$$\frac{\mathbb{E}_t[\|x_t^P - x_t\|]}{\mathbb{E}_t[\|x_t\|]} \leq \frac{\|\hat{c} - c\|}{\mathbb{E}_t[\|x_t\|]} + \|\hat{\Phi} - \Phi\|_2 \frac{\mathbb{E}_t[\|x_{t-1}\|]}{\mathbb{E}_t[\|x_t\|]} + \|\hat{\Theta} - \Theta\|_2 \frac{\mathbb{E}_t[\|u_{t-1}\|]}{\mathbb{E}_t[\|x_t\|]}.$$

Hence, using $\mathbb{E}_t[\|x_{t-1}\|] = \mathbb{E}_t[\|x_t\|]$ and the analogous relation for u_t, the relative expected norm of the prediction error is bounded by the sum of $c_{\text{err}}, \Phi_{\text{err}}, \Theta_{\text{err}}$ which are thus suitable measures of the errors of $\hat{c}$, $\hat{\Phi}$, and $\hat{\Theta}$, respectively.

- the same error measures – denoted by $c_{\text{ml}}^{\text{err}}$, $\Phi_{\text{ml}}^{\text{err}}$, and $\Theta_{\text{ml}}^{\text{err}}$ – for (approximations of) the maximum likelihood estimates. These estimates were obtained by two steps of Gauss-Newton optimization on the log-likelihood function with some basic step size control. The iteration was initialized with the moment-based estimates.

  These values are considered for comparison purposes only. A competitive ML method is outside the scope of this text.

- the average value of the relative perturbation

$$\Delta_{A,B} := \frac{\|\begin{bmatrix} A & B \end{bmatrix} - \begin{bmatrix} \tilde{A} & \tilde{B} \end{bmatrix}\|_F}{\|\begin{bmatrix} A & B \end{bmatrix}\|}$$

  resulting from solvability enforcement, with A and B as in Algorithm 2.

- a function of the form constant $\cdot N^{-1/2}$, to highlight the convergence rate of these quantities. Indeed, for Gaussian noise, the expected convergence rate is $N^{-1/2}$, as a consequence of the central limit theorem.

All given numbers are the averages over 10 different experiments (utilizing the same parameters c, Φ, Θ; but different noise vectors $\hat{u}_t$).

Experiment 1 The first model is of dimension $d = 4$ with

$$c = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \Phi = \begin{bmatrix} 0.16 & 0.20 & 0.12 & 0.09 \\ 0.13 & 0.03 & 0.10 & 0.02 \\ 0.20 & 0.15 & 0.12 & 0.16 \\ 0.16 & 0.06 & 0.19 & 0.08 \end{bmatrix}, \quad \Theta = \begin{bmatrix} 0.01 & -0.23 & 0.70 & -0.37 \\ 0.50 & 0 & 0.23 & 0.23 \\ -0.13 & -0.25 & -0.33 & -0.14 \\ -0.21 & 0.20 & -0.61 & 0.44 \end{bmatrix}. \tag{16}$$

The results are given in Figure 4. As one can see in the left plot, the finite-sample moments do converge to their asymptotic values with the expected rate of $N^{-1/2}$. Among the moments μ is estimated best, followed by M_0, M_1, and, finally, M_2. In this experiment we have $\rho(\Phi) \approx 0.5$, $\rho(\Theta) \approx 0.55$, i.e., the spectral radii are distant from 1, and thus solvability enforcement should not be needed for at least moderately accurate estimates of the moments. Indeed, as can be seen from the dotted curve depicting $\Delta_{A,B}$, solvability enforcement was not needed for $N > 4 \cdot 10^3$. When it was required, the perturbation on the moment estimates was negligible compared to the error on these values.

In the right plot we see that also the moment based parameter estimates converge with rate $N^{-1/2}$. The estimate of Θ is notably better than that of Φ with the error of c being somewhere in between those two. The errors of the maximum likelihood estimates, depicted in red, behave roughly like the moment-based estimates in terms of convergence rate. However, their absolute error is half an order of magnitude better.

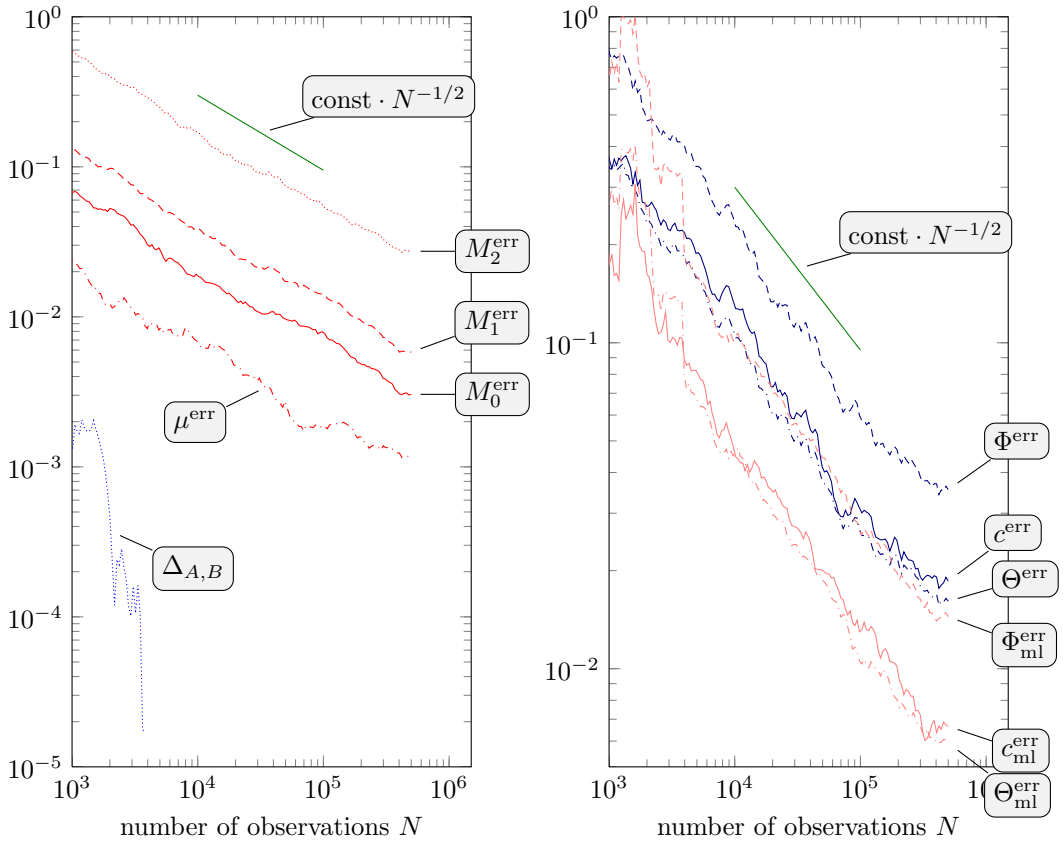


Figure 4: Development of the average error on the moment (left) and parameter (right) estimates in Experiment 1

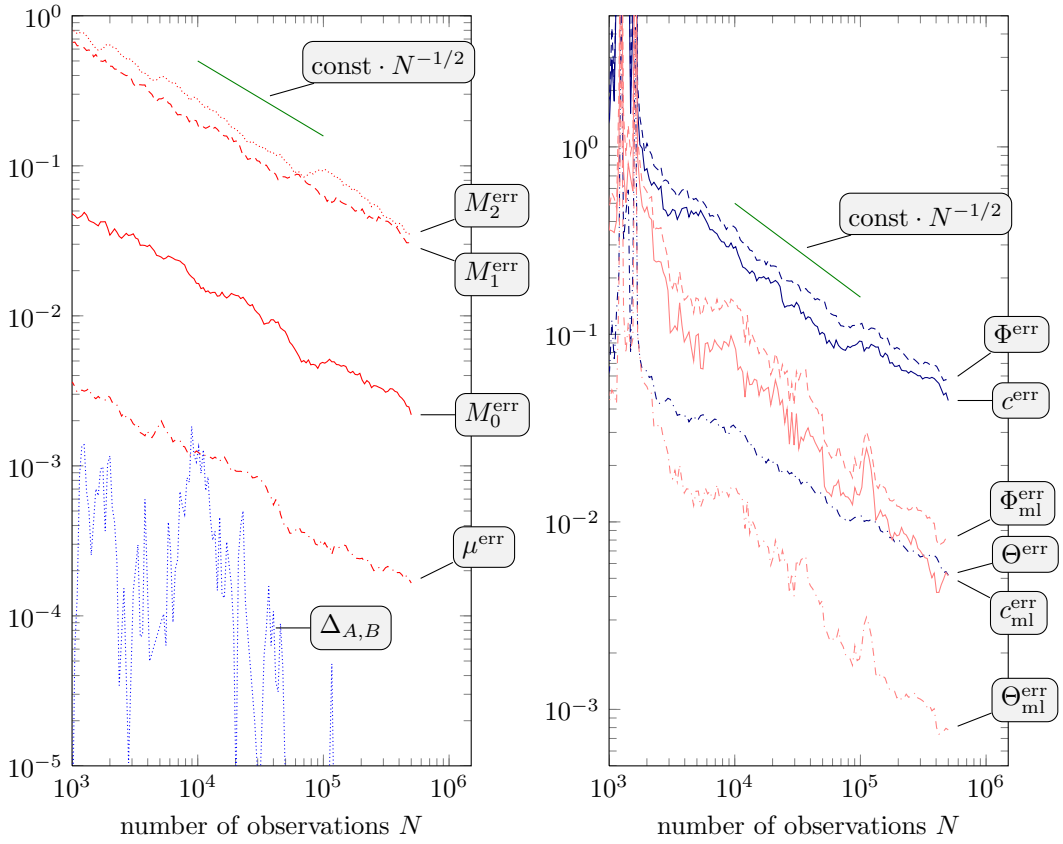


Figure 5: Development of the average error on the moment (left) and parameter (right) estimates in Experiment 2

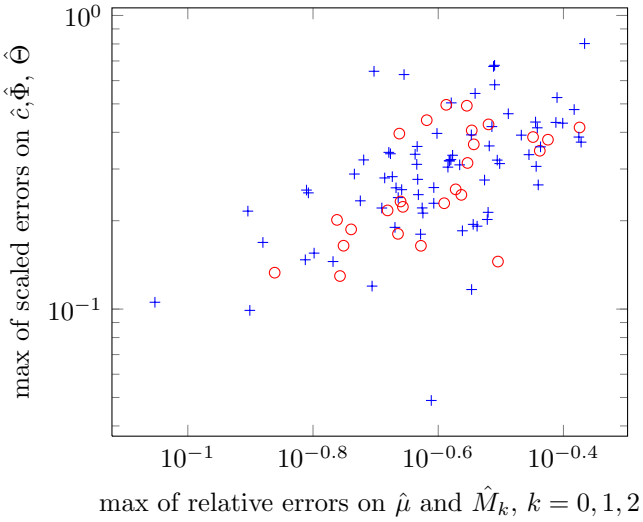


Figure 6: Accuracy of moments and estimates for 100 experiments of the model in (17), $N = 10000$

Experiment 2 The second model is given by

$$c = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \Phi = \begin{bmatrix} 0.84 & 0.084 \\ 0.042 & 0.84 \end{bmatrix}, \quad \Theta = \begin{bmatrix} 0.79 & 0.06 \\ 0.09 & 0.79 \end{bmatrix}. \tag{17}$$

The results are given in Figure 5. Almost all we said in the first example still applies here: the errors decline steadily; the method works.

The difference is that now the spectral radii are much closer to 1, namely $\rho(\Phi) \approx 0.9$ and $\rho(\Theta) \approx 0.86$. As a consequence, more observations are needed to get the same accuracy (although the convergence rate is still $N^{-1/2}$), and solvability enforcement of the matrix equation is needed until a much larger value of N. This example shows the importance of solvability enforcement for the robustness of the moment-based estimator.

As a further visualization of the impact of solvability enforcement, we set $N = 10^4$ and display a scatter plot (Figure 6) showing, for 100 different experiments, on the x-axis the maximum of the errors on the moments and on the y-axis the maximum of the errors on the moment-based parameter estimates. In some sense, the abscissa measures the error of the input data for our procedure, and the ordinate values measures the error of the output. The symbol used in the plot is a red $\circ$ for the experiments in which we needed to enforce solvability, and a blue $+$ if the resulting $Q(\lambda)$ was already positive definite on the whole unit circle.

One can see that the values line up diagonally, albeit very roughly. This is expected, since the larger the distance from the real moments, the worse the estimated parameters are, but this relation is not perfectly linear. Nevertheless, the red data points lie on average on this line rather than above it, and this supports the hypothesis that, when solvability enforcement is needed, it does not have a negative impact on the expected accuracy of the solution.

Moreover, in all these experiments, the cost of the solvability enforcement step was smaller than that of computing the finite-sample moments $\hat{M}_i$, since we are only performing computations with matrices of size $d \ll N$. On the other hand, even two steps of the maximum-likelihood estimator take significantly longer than the moment computation and the moment-based estimation procedure. For instance, on the second experiment, for $N = 5000$ the moment-based estimator needed only 2% of the CPU time used by ML.

We end this section with a remark on the seemingly large estimation error; 10^{-1} or 10^{-2} is usually not considered small in the linear algebra community. Nevertheless, this level

of accuracy is not uncommon in this kind of models, especially considering that we are estimating a large number of parameters using observations affected by large errors. The estimated data still show the main features of the underlying model.

6 Comments and conclusions

The availability of a fast and efficient moment based estimator is relevant in practical applications. The procedure we introduced can produce reasonably accurate parameter estimates in an efficient way. Its main shortcoming, the need to solve a possibly unsolvable matrix equation, is cured by our enforcement technique.

The results obtained with solvability enforcement are encouraging. The error obtained in the estimated model parameters after the regularization is of the same magnitude as that obtained in cases in which this technique is not needed. Hence, this additional step does not enlarge the error and makes the estimation procedure more resilient: in all cases where the original data results in an unsolvable model, with this technique we can return a meaningful result rather than aborting with a failure message. In some cases, this enforcement is needed in the majority of the experiments. In addition, positivity enforcement comes almost for free in terms of computational work, since the dominant cost within the moment-based estimator is the computation of the finite-sample moments (scaling with the length N of the time series), whereas all remaining steps only involve $d \times d$ (or $3d \times 3d$) matrices, independent of N.

Since the resulting estimator is fast, it can be used with more data, allowing for higher density in the time series (for stock exchange data, for instance, data at 5-minutes intervals are now commonly available).

Overall, our experiments seem to support the conclusions that:

- the moment-based estimator produces reasonably good and consistent parameter estimates;
- solvability enforcement makes the procedure much more robust at negligible cost;
- moment-based estimates are still inferior to ML estimates in terms of accuracy, but they are viable starting values for iterative maximum likelihood methods;
- the moment-based estimator is much faster than a ML procedure.

References

[1] R. Alam, S. Bora, M. Karow, V. Mehrmann, and J. Moro, *Perturbation theory for Hamiltonian matrices and the distance to bounded-realness*, SIAM J. Matrix Anal. Appl., 32 (2011), pp. 484–514.

[2] P. Benner and M. Voigt, *Spectral characterization and enforcement of negative imaginariness for descriptor systems*, Linear Algebra and its Applications, (2013), pp. –.

[3] T. Betcke and L. N. Trefethen, *Computations of eigenvalue avoidance in planar domains*, PAMM, Proceedings in Applied Mathematics and Mechanics, 4 (2004), pp. 634–635.

[4] S. Bora and V. Mehrmann, *Perturbation theory for structured matrix pencils arising in control theory*, SIAM J. Matrix Anal. Appl., 28 (2006), pp. 148–169.

[5] T. Brüll and C. Schröder, *Dissipativity enforcement via perturbation of para-hermitian pencils*, IEEE Transactions on Circuits and Systems I, 60 (2013), pp. 164–177.

[6] R. Byers, D. S. Mackey, V. Mehrmann, and H. Xu, *Symplectic, BVD, and palindromic approaches to discrete-time control problems*, MIMS EPrints 2008.35, The University of Manchester, 2008.

[7] J. G. De Gooijer and R. J. Hyndman, *25 years of time series forecasting*, International Journal of Forecasting, 22 (2006), pp. 443–473.

[8] J. C. Engwerda, A. C. M. Ran, and A. L. Rijkeboer, *Necessary and sufficient conditions for the existence of a positive definite solution of the matrix equation* $X + A^*X^{-1}A = Q$, Linear Algebra Appl., 186 (1993), pp. 255–275.

[9] I. Gohberg, P. Lancaster, and L. Rodman, *Matrix polynomials*, Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1982. Computer Science and Applied Mathematics.

[10] I. Gohberg, P. Lancaster, and L. Rodman, *Indefinite linear algebra and applications*, Birkhäuser Verlag, Basel, 2005.

[11] G. H. Golub and C. F. Van Loan, *Matrix computations*, Johns Hopkins Studies in the Mathematical Sciences, Johns Hopkins University Press, Baltimore, MD, third ed., 1996.

[12] S. Grivet-Talocia, *Passivity enforcement via perturbation of Hamiltonian matrices*, IEEE Trans. Circuits and Systems, 51 (2004), pp. 1755–1769.

[13] N. Guglielmi, D. Kressner, and C. Lubich, *Low-rank differential equations for hamiltonian matrix nearness problems*, technical report, Universität Tübingen, 2013.

[14] N. Guglielmi and M. L. Overton, *Fast algorithms for the approximation of the pseudospectral abscissa and pseudospectral radius of a matrix.* To appear in SIAM J. Matrix Anal. Appl.

[15] N. J. Higham and H.-M. Kim, *Numerical analysis of a quadratic matrix equation*, IMA J. Numer. Anal., 20 (2000), pp. 499–519.

[16] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, 1st ed., 1985.

[17] C. Kascha, *A comparison of estimation methods for vector autoregressive moving-average models*, Econometric Reviews, 31 (2012), pp. 297–324.

[18] T. Kato, *Perturbation theory for linear operators*, Classics in Mathematics, Springer-Verlag, Berlin, 1995. Reprint of the 1980 edition.

[19] D. Kressner, C. Schröder, and D. S. Watkins, *Implicit QR algorithms for palindromic and even eigenvalue problems*, Numer. Algorithms, 51 (2009), pp. 209–238.

[20] D. Kristensen and O. Linton, *A closed-form estimator for the GARCH* $(1,1)$ *model*, Econometric Theory, 22 (2006), pp. 323–337.

[21] P. D. Lax, *Linear Algebra and its Applications*, Pure and Applied Mathematics (Hoboken), Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, second ed., 2007.

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

[23] D. S. Mackey, N. Mackey, C. Mehl, and V. Mehrmann, *Structured polynomial eigenvalue problems: good vibrations from good linearizations*, SIAM J. Matrix Anal. Appl., 28 (2006), pp. 1029–1051 (electronic).

[24] L. Mátyás, ed., *Generalized Method of Moments Estimation*, Cambridge University Press, Cambridge, 1999.

[25] B. Meini, *Efficient computation of the extreme solutions of $X + A^*X^{-1}A = Q$ and $X - A^*X^{-1}A = Q$*, Math. Comp., 71 (2002), pp. 1189–1204 (electronic).

[26] G. Sbrana and F. Poloni, *A closed-form estimator for the multivariate GARCH(1,1) model*, Journal of Multivariate Analysis, 120 (2013), pp. 152–162.

[27] K. Schreiber, *Nonlinear Eigenvalue Problems: Newton-type Methods and Nonlinear Rayleigh Functionals*, PhD thesis, Technische Universität Berlin, 2008.

[28] C. Schröder, *Palindromic and Even Eigenvalue Problems - Analysis and Numerical Methods*, PhD thesis, Technical University Berlin, Germany, 2008.

[29] C. Schröder and T. Stykel, *Passivation of LTI systems*, Preprint 368, DFG Research Center Matheon, TU Berlin, 2007.

[30] M. Voigt and P. Benner, *Passivity enforcement of descriptor systems via structured perturbation of hamiltonian matrix pencils*. Talk at Meeting of the GAMM Activity Group Dynamics and Control Theory, Linz, 2011; paper in preparation.

[31] Y. Wang, Z. Zhang, C.-K. Koh, G. Pang, and N. Wong, *PEDS: Passivity enforcement for descriptor systems via hamiltonian-symplectic matrix pencil perturbation*, in Computer-Aided Design (ICCAD), 2010 IEEE/ACM International Conference on, nov. 2010, pp. 800 –807.