

# A Random Matrix Approach to VARMA Processes

Zdzisław Burda,<sup>1,\*</sup> Andrzej Jarosz,<sup>2,†</sup> Maciej A. Nowak,<sup>1,‡</sup> and Małgorzata Snarska<sup>1,3,§</sup>

<sup>1</sup>*Marian Smoluchowski Institute of Physics and Mark Kac Complex Systems Research Centre,  
Jagiellonian University, Reymonta 4, 30-059 Kraków, Poland*

<sup>2</sup>*Clico Ltd., Oleandry 2, 30-063 Kraków, Poland*

<sup>3</sup>*Cracow University of Economics, Department of Econometrics  
and Operations Research, Rakowicka 27, 31-510 Kraków, Poland*

(Dated: February 4, 2010)

We apply random matrix theory to derive spectral density of large sample covariance matrices generated by multivariate VMA( $q$ ), VAR( $q$ ) and VARMA( $q_1, q_2$ ) processes. In particular, we consider a limit where the number of random variables  $N$  and the number of consecutive time measurements  $T$  are large but the ratio  $N/T$  is fixed. In this regime the underlying random matrices are asymptotically equivalent to Free Random Variables (FRV). We apply the FRV calculus to calculate the eigenvalue density of the sample covariance for several VARMA-type processes. We explicitly solve the VARMA(1,1) case and demonstrate a perfect agreement between the analytical result and the spectra obtained by Monte Carlo simulations. The proposed method is purely algebraic and can be easily generalized to  $q_1 > 1$  and  $q_2 > 1$ .

PACS numbers: 89.65.Gh (Economics; econophysics, financial markets, business and management), 02.50.Sk (Multivariate analysis), 02.60.Cb (Numerical simulation; solution of equations), 02.70.Uu (Applications of Monte Carlo methods)

Keywords: VARMA, random matrix theory, free random variables, Wishart ensemble, covariance matrix, historical estimation

## I. INTRODUCTION

Vector auto-regressive (VAR) models play an important role in contemporary macro-economics, being an example of an approach called the “dynamic stochastic general equilibrium” (DSGE), which is superseding traditional large-scale macro-econometric forecasting methodologies [1]. The motivation behind them is based on the assertion that more recent values of a variable are more likely to contain useful information about its future movements than the older ones. On the other hand, a standard tool in multivariate time series analysis is vector moving average (VMA) models, which is really a linear regression of the present value of the time series w.r.t. the past values of a white noise. A broader class of stochastic processes used in macro-economy comprises both these kinds together in the form of vector auto-regressive moving average (VARMA) models. These methodologies can capture certain spatial and temporal structures of multidimensional variables which are often neglected in practice; including them not only results in more accurate estimation, but also leads to models which are more interpretable. They are widely used by academia and central banks (*cf.* the European Central Bank’s Smets–Wouters model for the euro zone [2]), as they constitute quite a simple version of the DSGE equations.

VARMA models are constructed from a number of univariate ARMA (Box–Jenkins; see for example [3]) processes, typically coupled with each other. In this paper, we investigate only a significantly simplified circumstance when there is no coupling between the many ARMA components. One may argue that this is too far fetched and will be of no use in describing an economic reality. However, one may also treat it as a “zeroth-order hypothesis,” analogously to the idea of [4, 5] in finance, namely that the case with no cross-covariances is considered theoretically, and subsequently compared to some real-world data modeled by a VARMA process; any discrepancy between the two will reflect

---

\*Electronic address: [zdzislaw.burda@uj.edu.pl](mailto:zdzislaw.burda@uj.edu.pl)

†Electronic address: [andrzej.jarosz@clico.pl](mailto:andrzej.jarosz@clico.pl)

‡Electronic address: [nowak@th.if.uj.edu.pl](mailto:nowak@th.if.uj.edu.pl)

§Electronic address: [snarska@th.if.uj.edu.pl](mailto:snarska@th.if.uj.edu.pl)

nontrivial cross-covariances present in the system, thus permitting their investigation. This latter route is taken in this communication.

A challenging and yet increasingly important problem is the estimation of large covariance matrices generated by these stationary VARMA( $q_1, q_2$ ) processes, since high dimensionality of the data as compared to the sample size is quite common in many statistical problems (the “dimensionality curse”). Therefore, an appropriate “noise cleaning” procedure has to be implemented, and random matrix theory (RMT) provides a natural and efficient outfit for doing that. In particular, the mean spectral densities (a.k.a. “limiting spectral distributions,” LSD) of the Pearson estimators of the cross-covariances for the VMA(1) and VAR(1) models, in the relevant high-dimensionality sector and under the full decoupling, have been derived in [6] by applying the framework proposed by [7].

In this paper, we suggest that such calculations can be considerably simplified by resorting to a mathematical concept of the free random variables (FRV) calculus [8, 9], succinctly introduced in sec. II. Our general FRV formula [10] allows not only to rediscover, which much less strain, the two fourth-order polynomial equations obtained in [6] in the VMA(1) and VAR(1) cases, but also to derive a sixth-order equation (45) which produces the mean spectral density for a more involved VARMA(1, 1) model. The results are verified by numerical simulations, which show a perfect agreement. This is all done in sec. III.

## II. DOUBLY CORRELATED WISHART ENSEMBLES AND FREE RANDOM VARIABLES

### A. Doubly Correlated Wishart Ensembles

#### 1. Correlated Gaussian Random Variables

VARMA( $q_1, q_2$ ) stochastic processes, as we will see below, fall within quite a general set-up encountered in many areas of science where a probabilistic nature of multiple degrees of freedom evolving in time is relevant, for example, multivariate time series analysis in finance, applied macro-econometrics and engineering. To describe this framework, consider a situation of  $N$  time-dependent random variables which are measured at  $T$  consecutive time moments (separated by some time interval  $\delta t$ ); let  $Y_{ia}$  be the value of the  $i$ -th ( $i = 1, \dots, N$ ) random number at the  $a$ -th time moment ( $a = 1, \dots, T$ ); together, they make up a rectangular  $N \times T$  matrix  $\mathbf{Y}$ . In what usually would be the first approximation, each  $Y_{ia}$  is supposed to be drawn from a Gaussian probability distribution. We will also assume that they have mean values zero,  $\langle Y_{ia} \rangle = 0$ . These degrees of freedom may in principle display mutual correlations. A set of correlated zero-mean Gaussian numbers is fully characterized by the two-point covariance function,  $\mathcal{C}_{ia,jb} \equiv \langle Y_{ia} Y_{jb} \rangle$  if the underlying stochastic process generating these numbers is stationary. Linear stochastic processes, including VARMA( $q_1, q_2$ ), belong to this category. We will restrict our attention to an even narrower class where the cross-correlations between different variables and the auto-correlations between different time moments are factorized, *i.e.*,

$$\langle Y_{ia} Y_{jb} \rangle = C_{ij} A_{ab}. \quad (1)$$

In this setting, the inter-variable covariances do not change in time (and are described by an  $N \times N$  cross-covariance matrix  $\mathbf{C}$ ), and also the temporal covariances are identical for all the numbers (and are included in a  $T \times T$  auto-covariance matrix  $\mathbf{A}$ ; both these matrices are symmetric and positive-definite). The Gaussian probability measure with this structure of covariances is known from textbooks,

$$\begin{aligned} P_{\text{c.G.}}(\mathbf{Y}) D\mathbf{Y} &= \frac{1}{\mathcal{N}_{\text{c.G.}}} \exp \left( -\frac{1}{2} \sum_{i,j=1}^N \sum_{a,b=1}^T Y_{ia} [\mathbf{C}^{-1}]_{ij} Y_{jb} [\mathbf{A}^{-1}]_{ba} \right) D\mathbf{Y} = \\ &= \frac{1}{\mathcal{N}_{\text{c.G.}}} \exp \left( -\frac{1}{2} \text{Tr} \mathbf{Y}^T \mathbf{C}^{-1} \mathbf{Y} \mathbf{A}^{-1} \right) D\mathbf{Y}, \end{aligned} \quad (2)$$

where the normalization constant  $\mathcal{N}_{\text{c.G.}} = (2\pi)^{NT/2} (\text{Det} \mathbf{C})^{T/2} (\text{Det} \mathbf{A})^{N/2}$ , and the integration measure  $D\mathbf{Y} \equiv \prod_{i=1}^N \prod_{a=1}^T dY_{ia}$ , while the letters “c.G.” stand for “correlated Gaussian.”

Now, a standard way to approach correlated Gaussian random numbers is to recall that they can always be decomposed as linear combinations of uncorrelated Gaussian degrees of freedom; indeed, this is achieved through the transformation

$$\mathbf{Y} = \sqrt{\mathbf{C}}\tilde{\mathbf{Y}}\sqrt{\mathbf{A}}, \quad \text{which yields} \quad P_{\mathbf{G}}(\tilde{\mathbf{Y}})D\tilde{\mathbf{Y}} = \frac{1}{\mathcal{N}_{\mathbf{G}}} \exp\left(-\frac{1}{2}\text{Tr}\tilde{\mathbf{Y}}^T\tilde{\mathbf{Y}}\right)D\tilde{\mathbf{Y}}, \quad (3)$$

where the square roots of the covariance matrices, necessary to facilitate the transition, exist due to the positive-definiteness of  $\mathbf{C}$  and  $\mathbf{A}$ ; the new normalization reads  $\mathcal{N}_{\mathbf{G}} = (2\pi)^{NT/2}$ .

## 2. Estimating Equal-Time Cross-Covariances

An essential problem in multivariate analysis is to determine (estimate) the covariance matrices  $\mathbf{C}$  and  $\mathbf{A}$  from given  $N$  time series of length  $T$  of the realizations of our random variables  $Y_{ia}$ . For simplicity, we do not distinguish in notation between random numbers, *i.e.*, the population, and their realizations in actual experiments, *i.e.*, the sample. Since the realized cross-covariance between degrees  $i$  and  $j$  at the same time  $a$  is  $Y_{ia}Y_{ja}$ , the simplest method to estimate the today's cross-covariance  $c_{ij}$  is to compute the time average,

$$c_{ij} \equiv \frac{1}{T} \sum_{a=1}^T Y_{ia}Y_{ja}, \quad \text{i.e.,} \quad \mathbf{c} = \frac{1}{T}\mathbf{Y}\mathbf{Y}^T = \frac{1}{T}\sqrt{\mathbf{C}}\tilde{\mathbf{Y}}\mathbf{A}\tilde{\mathbf{Y}}^T\sqrt{\mathbf{C}}. \quad (4)$$

This is usually named the ‘‘Pearson estimator’’, up to the prefactor which depending on the context is  $1/(T-1)$  or  $1/T$ . Other estimators might be introduced, such as between distinct degrees of freedom at separate time moments (‘‘time-delayed estimators’’), or with certain decreasing weights given to older measurements to reflect their growing obsolescence (‘‘weighted estimators’’), but we will not investigate them in this article. Furthermore, in the last equality in (4), we cast  $\mathbf{c}$  through the uncorrelated Gaussian numbers contained in  $\tilde{\mathbf{Y}}$ , the price to pay for this being that the covariance matrices now enter into the expression for  $\mathbf{c}$ , making it more complicated; this will be the form used hereafter. The random matrix  $\mathbf{c}$  is called a ‘‘doubly correlated Wishart ensemble’’ [11].

Let us also mention that the auto-covariance matrix  $\mathbf{A}$  can be estimated through  $\mathbf{a} \equiv (1/N)\mathbf{Y}^T\mathbf{Y}$ . However, it is verified that this object carries identical information to the one contained in  $\mathbf{c}$  (it is ‘‘dual’’ to  $\mathbf{c}$ ), and therefore may safely be discarded. Indeed, these two estimators have same non-zero eigenvalues (modulo an overall rescaling by  $r$ ), and the larger one has  $|T-N|$  additional zero modes.

Any historical estimator is inevitably marred by the measurement noise; it will reflect the true covariances only to a certain degree, with a superimposed broadening due to the finiteness of the time series. More precisely, there are  $N(N+1)/2$  independent elements in  $\mathbf{C}$ , to be estimated from  $NT$  measured quantities  $\mathbf{Y}$ , hence the estimation accuracy will depend on the ‘‘rectangularity ratio,’’

$$r \equiv \frac{N}{T}; \quad (5)$$

the closer  $r$  to zero, the more truthful the estimate. This is a cornerstone of classical multivariate analysis. Unfortunately, a practical situation will typically feature a large number of variables sampled over a comparably big number of time snapshots, so that we may approximately talk about the ‘‘thermodynamical limit,’’

$$N \rightarrow \infty, \quad T \rightarrow \infty, \quad \text{such that} \quad r = \text{fixed}. \quad (6)$$

On the other hand, it is exactly this limit in which the FRV calculus (see the subsection below for its brief elucidation) can be applied; hence, the challenge of de-noising is somewhat counterbalanced by the computationally powerful FRV techniques.

## B. A Short Introduction to the Free Random Variables Calculus: The Multiplication Algorithm

### 1. The $M$ -Transform and the Spectral Density

Any study of a (real symmetric  $K \times K$ ) random matrix  $\mathbf{H}$  will most surely include a fundamental question about the average values of its (real) eigenvalues  $\lambda_1, \dots, \lambda_K$ . They are concisely encoded in the “mean spectral density,”

$$\rho_{\mathbf{H}}(\lambda) \equiv \frac{1}{K} \sum_{i=1}^K \langle \delta(\lambda - \lambda_i) \rangle = \frac{1}{K} \langle \text{Tr}(\lambda \mathbf{1}_K - \mathbf{H}) \rangle. \quad (7)$$

Here the expectation map  $\langle \dots \rangle$  is understood to be taken w.r.t. the probability measure  $P(\mathbf{H})D\mathbf{H}$  of the random matrix. We will always have this distribution rotationally (*i.e.*,  $\mathbf{H} \rightarrow \mathbf{O}^T \mathbf{H} \mathbf{O}$ , with  $\mathbf{O}$  orthogonal) invariant, and hence the full information about  $\mathbf{H}$  resides in its eigenvalues, distributed on average according to (7).

On the practical side, it is more convenient to work with either of the two equivalent objects,

$$G_{\mathbf{H}}(z) \equiv \frac{1}{K} \left\langle \text{Tr} \frac{1}{z \mathbf{1}_K - \mathbf{H}} \right\rangle, \quad \text{or} \quad M_{\mathbf{H}}(z) \equiv z G_{\mathbf{H}}(z) - 1, \quad (8)$$

referred to as the “Green’s function” (or the “resolvent”) and the “ $M$ -transform” of  $\mathbf{H}$ . The latter is also called the “moments’ generating function,” since if the “moments”  $M_{\mathbf{H},n} \equiv (1/K) \langle \text{Tr} \mathbf{H}^n \rangle$  of  $\mathbf{H}$  exist, it can be expanded into a power series around  $z \rightarrow \infty$  as  $M_{\mathbf{H}}(z) = \sum_{n \geq 1} M_{\mathbf{H},n} / z^n$ . It should however be underlined that even for probability measures disallowing such an expansion (heavy-tailed distributions, preeminent in finance, being an example), the quantities (8) still manage to entirely capture the spectral properties of  $\mathbf{H}$ ; hence the name “ $M$ -transform” more appropriate, in addition to being more compact.

We will show that for our purposes (multiplication of random matrices; see par. IIB2) the  $M$ -transform serves better than the Green’s function. However, it is customary to write the relationship between (7) and (8) in terms of this latter,

$$\rho_{\mathbf{H}}(\lambda) = -\frac{1}{\pi} \lim_{\epsilon \rightarrow 0^+} \text{Im} G_{\mathbf{H}}(\lambda + i\epsilon) = -\frac{1}{2\pi i} \lim_{\epsilon \rightarrow 0^+} (G_{\mathbf{H}}(\lambda + i\epsilon) - G_{\mathbf{H}}(\lambda - i\epsilon)). \quad (9)$$

resulting from a well-known formula for generalized functions,  $\lim_{\epsilon \rightarrow 0^+} 1/(x \pm i\epsilon) = \text{pv}(1/x) \mp i\pi\delta(x)$ .

### 2. The $N$ -Transform and Free Random Variables

The doubly correlated Wishart ensemble  $\mathbf{c}$  (4) may be viewed as a product of several random and non-random matrices. The general problem of multiplying random matrices seems formidable. In classical probability theory, it can be effectively handled in the special situation when the random terms are independent: then, the exponential map reduces it to the addition problem of independent random numbers, solved by considering the logarithm of the characteristic functions of the respective PDFs, which proves to be additive. In matrix probability theory, a crucial insight came from D. Voiculescu and coworkers and R. Speicher [8, 9], who showed how to parallel the commutative construction in the noncommutative world. It starts with the notion of “freeness,” which basically comprises probabilistic independence together with a lack of any directional correlation between two random matrices. This nontrivial new property happens to be the right extension of classical independence, as it allows for an efficient algorithm of multiplying free random variables (FRV), which we state below:

**Step 1:** Suppose we have two random matrices,  $\mathbf{H}_1$  and  $\mathbf{H}_2$ , mutually free. Their spectral properties are best wrought into the  $M$ -transforms (8),  $M_{\mathbf{H}_1}(z)$  and  $M_{\mathbf{H}_2}(z)$ .

**Step 2:** The critical maneuver is to turn attention to the functional inverses of these  $M$ -transforms, the so-called “ $N$ -transforms,”

$$M_{\mathbf{H}}(N_{\mathbf{H}}(z)) = N_{\mathbf{H}}(M_{\mathbf{H}}(z)) = z. \quad (10)$$

**Step 3:** The  $N$ -transforms submit to a very straightforward rule upon multiplying free random matrices (the “FRV multiplication law”),

$$N_{\mathbf{H}_1\mathbf{H}_2}(z) = \frac{z}{1+z} N_{\mathbf{H}_1}(z) N_{\mathbf{H}_2}(z), \quad \text{for free } \mathbf{H}_1, \mathbf{H}_2. \quad (11)$$

**Step 4:** Finally, it remains to functionally invert the resulting  $N$ -transform  $N_{\mathbf{H}_1\mathbf{H}_2}(z)$  to gain the  $M$ -transform of the product,  $M_{\mathbf{H}_1\mathbf{H}_2}(z)$ , and consequently, all the spectral properties via formula (9).

It is stunning that such a simple prescription (relying on the choice of the  $M$ -transform as the carrier of the mean spectral information, and the construction of its functional inverse, the  $N$ -transform, which essentially multiplies under taking the free product) resolves the multiplication problem for free random noncommutative objects.

Let us just mention that the addition problem may be tackled along similar lines: In this case, the Green’s function should be exploited, its functional inverse considered ( $G_{\mathbf{H}}(B_{\mathbf{H}}(z)) = B_{\mathbf{H}}(G_{\mathbf{H}}(z)) = z$ ; it is sometimes called the “Blue’s function” [12, 13]), which obeys the “FRV addition law,”  $B_{\mathbf{H}_1+\mathbf{H}_2}(z) = B_{\mathbf{H}_1}(z) + B_{\mathbf{H}_2}(z) - 1/z$ , for two free random matrices. In this paper, we do not resort to using this addition formula, even though our problem could be approached through it as well.

Let us also remark that in the original mathematical formulations [8, 9] of these frames, a slightly different language is employed: Instead of the  $N$ -transform, the “ $S$ -transform” is found convenient,  $S_{\mathbf{H}}(z) \equiv (1+z)/(zN_{\mathbf{H}}(z))$ , while in place of the Blue’s function, one engages the “ $R$ -transform,”  $R_{\mathbf{H}}(z) \equiv B_{\mathbf{H}}(z) - 1/z$ . They fulfil simpler laws,  $S_{\mathbf{H}_1\mathbf{H}_2}(z) = S_{\mathbf{H}_1}(z)S_{\mathbf{H}_2}(z)$  and  $R_{\mathbf{H}_1+\mathbf{H}_2}(z) = R_{\mathbf{H}_1}(z) + R_{\mathbf{H}_2}(z)$ , respectively.

### 3. Doubly Correlated Wishart Ensembles from Free Random Variables

The innate potential of the FRV multiplication algorithm (11) is surely revealed when inspecting the doubly correlated Wishart random matrix  $\mathbf{c} = (1/T)\sqrt{\tilde{\mathbf{C}}\tilde{\mathbf{Y}}\mathbf{A}\tilde{\mathbf{Y}}^T\sqrt{\tilde{\mathbf{C}}}$  (4). This has been done in detail in [10], so we will only accentuate the main results here, referring the reader to the original paper for a thorough explanation.

The idea is that one uses twice the cyclic property of the trace (which permits cyclic shifts in the order of the terms), and twice the FRV multiplication law (11) (to break the  $N$ -transforms of products of matrices down to their constituents), in order to reduce the problem to solving the uncorrelated Wishart ensemble  $(1/T)\tilde{\mathbf{Y}}^T\tilde{\mathbf{Y}}$ . This last model is further simplified, again by the cyclic property and the FRV multiplication rule applied once, to the standard **GOE** random matrix squared (and the projector  $\mathbf{P} \equiv \text{diag}(\mathbf{1}_N, \mathbf{0}_{T-N})$ , designed to chip the rectangle  $\tilde{\mathbf{Y}}$  off the square **GOE**), whose properties are firmly established. Let us sketch the derivation,

$$\begin{aligned} N_{\mathbf{c}}(z) &\stackrel{\text{cyclic}}{\downarrow} N_{\frac{1}{T}\tilde{\mathbf{Y}}\mathbf{A}\tilde{\mathbf{Y}}^T\mathbf{C}}(z) \stackrel{\text{FRV}}{\downarrow} \frac{z}{1+z} N_{\frac{1}{T}\tilde{\mathbf{Y}}\mathbf{A}\tilde{\mathbf{Y}}^T}(z) N_{\mathbf{C}}(z) \stackrel{\text{cyclic}}{\downarrow} \\ &\stackrel{\text{cyclic}}{\downarrow} \frac{z}{1+z} N_{\frac{1}{T}\tilde{\mathbf{Y}}^T\tilde{\mathbf{Y}}\mathbf{A}}(rz) N_{\mathbf{C}}(z) \stackrel{\text{FRV}}{\downarrow} \frac{z}{1+z} \frac{rz}{1+rz} N_{\frac{1}{T}\tilde{\mathbf{Y}}^T\tilde{\mathbf{Y}}}(rz) N_{\mathbf{A}}(rz) N_{\mathbf{C}}(z) = \\ &= rz N_{\mathbf{A}}(rz) N_{\mathbf{C}}(z). \end{aligned} \quad (12)$$

This is the basic formula. Since the spectral properties of  $\mathbf{c}$  are given by its  $M$ -transform,  $M \equiv M_{\mathbf{c}}(z)$ , it is more pedagogical to recast (12) as an equation for the unknown  $M$ ,

$$z = rMN_{\mathbf{A}}(rM)N_{\mathbf{C}}(M). \quad (13)$$

It provides a means for computing the mean spectral density of a doubly correlated Wishart random matrix once the “true” covariance matrices  $\mathbf{C}$  and  $\mathbf{A}$  are given.

In this communication, only a particular instance of this fundamental formula is applied, namely with an arbitrary auto-covariance matrix  $\mathbf{A}$ , but with trivial cross-covariances,  $\mathbf{C} = \mathbf{1}_N$ . Using that  $N_{\mathbf{1}_K}(z) = 1 + 1/z$ , equation (13) thins out to

$$rM = M_{\mathbf{A}}\left(\frac{z}{r(1+M)}\right), \quad (14)$$

which will be strongly exploited below. Let us mention that these equalities (13), (14) have been derived through other, more tedious, techniques (the planar Feynman-diagrammatic expansion, the replica trick) in [14–18].

### III. VARMA FROM FREE RANDOM VARIABLES

In what follows, we will assume that the  $\text{VMA}(q)$ ,  $\text{VAR}(q)$ , or  $\text{VARMA}(q_1, q_2)$  stochastic processes are covariance (weak) stationary; for details, we refer to [19]. It implies certain restrictions on their parameters, but we will not bother with this issue in the current work. Another consequence is that the processes display some interesting features, such as invertibility.

For all this, we must in particular take both  $N$  and  $T$  large from the start, with their ratio  $r \equiv N/T$  fixed (6). More precisely, we stretch the range of the  $a$ -index from minus to plus infinity. This means that all the finite-size effects (appearing at the ends of the time series) are readily disregarded. In particular, there is no need to care about initial conditions for the processes, and all the recurrence relations are assumed to continue to the infinite past.

#### A. The $\text{VMA}(q)$ Process

##### 1. The Definition of $\text{VMA}(q)$

We consider a situation when  $N$  stochastic variables evolve according to identical independent  $\text{VMA}(q)$  (vector moving average) processes, which we sample over a time span of  $T$  moments. This is a simple generalization of the standard univariate weak-stationary moving average  $\text{MA}(q)$ . In such a setting, the value  $Y_{ia}$  of the  $i$ -th ( $i = 1, \dots, N$ ) random variable at time moment  $a$  ( $a = 1, \dots, T$ ) can be expressed as

$$Y_{ia} = \sum_{\alpha=0}^q a_{\alpha} \epsilon_{i, a-\alpha}. \quad (15)$$

Here all the  $\epsilon_{ia}$ 's are IID standard (mean zero, variance one) Gaussian random numbers (white noise),  $\langle \epsilon_{ia} \epsilon_{jb} \rangle = \delta_{ij} \delta_{ab}$ . The  $a_{\alpha}$ 's are some  $(q+1)$  real constants; importantly, they do not depend on the index  $i$ , which reflects the fact that the processes are identical and independent (no ‘‘spatial’’ covariances among the variables). The rank  $q$  of the process is a positive integer.

##### 2. The Auto-Covariance Matrix

In order to handle such a process (15), notice that the  $Y_{ia}$ 's, being linear combinations of uncorrelated Gaussian numbers, must also be Gaussian random variables, albeit displaying some correlations. Therefore, to fully characterize these variables, it is sufficient to calculate their two-point covariance function; this is straightforwardly done (see appendix 1 for details),

$$\langle Y_{ia} Y_{jb} \rangle = \delta_{ij} A_{ab}^{(1)}, \quad (16)$$

where

$$A_{ab}^{(1)} = \kappa_0^{(1)} \delta_{ab} + \sum_{d=1}^q \kappa_d^{(1)} (\delta_{a, b-d} + \delta_{a, b+d}), \quad \text{with} \quad \kappa_d^{(1)} \equiv \sum_{\alpha=0}^{q-d} a_{\alpha} a_{\alpha+d}, \quad d = 0, 1, \dots, q. \quad (17)$$

In other words, the cross-covariance matrix is trivial,  $\mathbf{C} = \mathbf{1}_N$  (no correlations between different variables), while the auto-covariance matrix  $\mathbf{A}^{(1)}$ , responsible for temporal correlations, can be called ‘‘ $(2q+1)$ -diagonal.’’ In the course of this article, we will use several different auto-covariance matrices, and for brevity, we decide to label them with superscripts; their definitions are all collected in appendix 2.

For example, in the simplest case of  $\text{VMA}(1)$ , it is tri-diagonal,

$$A_{ab}^{(1)} = (a_0^2 + a_1^2) \delta_{ab} + a_0 a_1 (\delta_{a, b-1} + \delta_{a, b+1}). \quad (18)$$

### 3. The Fourier Transform and the $M$ -Transform of the Auto-Covariance Matrix

Such an infinite matrix (17) is translationally invariant (as announced, it is one of the implications of the weak stationarity), *i.e.*, the value of any of its entries depends only on the distance between its indices,  $A_{ab}^{(1)} = A^{(1)}(a - b)$ ; specifically,  $A^{(1)}(\pm d) = \kappa_d^{(1)}$ , for  $d = 0, 1, \dots, q$ , and  $A^{(1)}(|d| > q) = 0$ . Hence, it is convenient to rewrite this matrix in the Fourier space,

$$\widehat{A}^{(1)}(p) \equiv \sum_{d \in \mathbb{Z}} e^{idp} A^{(1)}(d) = \kappa_0^{(1)} + 2 \sum_{d=1}^q \kappa_d^{(1)} \cos(dp). \quad (19)$$

In this representation, the  $M$ -transform of  $\mathbf{A}^{(1)}$  is readily obtained [10],

$$M_{\mathbf{A}^{(1)}}(z) = \frac{1}{2\pi} \int_{-\pi}^{\pi} dp \frac{\widehat{A}^{(1)}(p)}{z - \widehat{A}^{(1)}(p)}. \quad (20)$$

This integral can be evaluated by the method of residues for any value of  $q$ , which we do in appendix 3, where also we print the general result (42). In particular, for  $q = 1$ ,

$$M_{\mathbf{A}^{(1)}}(z) = \frac{z}{\sqrt{z - (a_0 + a_1)^2} \sqrt{z - (a_0 - a_1)^2}} - 1, \quad (21)$$

where the square roots are principal.

### 4. The Pearson Estimator of the Covariances from Free Random Variables

We will be interested in investigating the spectral properties of the Pearson estimator  $\mathbf{c} = (1/T)\mathbf{Y}\mathbf{Y}^T = (1/T)\widetilde{\mathbf{Y}}\mathbf{A}^{(1)}\widetilde{\mathbf{Y}}^T$  (4). The  $M$ -transform of this correlated Wishart random matrix,  $M \equiv M_{\mathbf{c}}(z)$ , can be retrieved from equation (14). We could write it for any  $q$  using (42), but we will restrict ourselves only to  $q = 1$ , in which case the substitution of (21) leads to a fourth-order polynomial (Ferrari) equation for the unknown  $M$ ,

$$\begin{aligned} & r^4 (a_0^2 - a_1^2)^2 M^4 + 2r^3 \left( - (a_0^2 + a_1^2) z + (a_0^2 - a_1^2)^2 (r + 1) \right) M^3 + \\ & + r^2 \left( z^2 - 2 (a_0^2 + a_1^2) (r + 2) z + (a_0^2 - a_1^2)^2 (r^2 + 4r + 1) \right) M^2 + \\ & + 2r \left( z^2 - (a_0^2 + a_1^2) (2r + 1) z + (a_0^2 - a_1^2)^2 r (r + 1) \right) M + r \left( -2 (a_0^2 + a_1^2) z + (a_0^2 - a_1^2)^2 r \right) = 0. \end{aligned} \quad (22)$$

The FRV technique allowed us therefore to find this equation in a matter of a few lines of a simple algebraic computation. It has already been derived in [6], and (22) may be verified to coincide with the version given in that paper. In [6], the pertinent equation is printed before (A.6), and to compare the two, one needs to change their variables into ours according to  $y \rightarrow 1/r$ ,  $x \rightarrow z/r$ , and  $\underline{m} \rightarrow -r(1 + M)/z$ . The last equality means that  $\underline{m}$  and  $m$  of [6] correspond in our language to the Green's functions  $-rG_{\mathbf{c}}(z)$  and  $-G_{\mathbf{a}}(z/r)$ , respectively, where  $\mathbf{a} = (1/N)\mathbf{Y}^T\mathbf{Y}$  is the Pearson estimator dual to  $\mathbf{c}$ . As mentioned, a quick extension to the case of arbitrary  $q$  is possible, however the resulting equations for  $M$  will be significantly more complicated; for instance, for  $q = 2$ , a lengthy ninth-order polynomial equation is discovered.

## B. The VAR( $q$ ) Process

### 1. The Definition of VAR( $q$ )

A set-up of  $N$  identical and independent VAR( $q$ ) (vector auto-regressive) processes is somewhat akin to (15), *i.e.*, we consider  $N$  decoupled copies of a standard univariate AR( $q$ ) process,

$$Y_{ia} - \sum_{\beta=1}^q b_{\beta} Y_{i,a-\beta} = a_0 \epsilon_{ia}. \quad (23)$$

It is again described by the demeaned and standardized Gaussian white noise  $\epsilon_{ia}$  (which triggers the stochastic evolution), as well as  $(q+1)$  real constants  $a_0, b_{\beta}$ , with  $\beta = 1, \dots, q$ . As announced before, the time stretches to the past infinity, so no initial condition is necessary. Although at first sight (23) may appear to be a more involved recurrence relation for the  $Y_{ia}$ 's, it is actually easily reduced to the VMA( $q$ ) case: It remains to remark that if one exchanges the  $Y_{ia}$ 's with the  $\epsilon_{ia}$ 's, one precisely arrives at the VMA( $q$ ) process with the constants  $a_0^{(2)} \equiv 1/a_0$ ,  $a_{\beta}^{(2)} \equiv -b_{\beta}/a_0$ ,  $\beta = 1, \dots, q$ . In other words, the auto-covariance matrix  $\mathbf{A}^{(3)}$  of the VAR( $q$ ) process (23) is simply the inverse of the auto-covariance matrix  $\mathbf{A}^{(2)}$  of the corresponding VMA( $q$ ) process with the described modification of the parameters,

$$\mathbf{A}^{(3)} = \left( \mathbf{A}^{(2)} \right)^{-1}. \quad (24)$$

This inverse exists thanks to the weak stationarity supposition.

### 2. The Fourier Transform and the $M$ -Transform of the Auto-Covariance Matrix

The Fourier transform of the auto-covariance matrix  $\mathbf{A}^{(3)}$  of VAR( $q$ ) is therefore a (number) inverse of its counterpart for VMA( $q$ ) with its parameters appropriately changed,

$$\widehat{A}^{(3)}(p) = \frac{1}{\widehat{A}^{(2)}(p)} = \frac{1}{\kappa_0^{(2)} + 2 \sum_{d=1}^q \kappa_d^{(2)} \cos(dp)}, \quad (25)$$

where

$$\kappa_d^{(2)} = \frac{1}{a_0^2} \sum_{\alpha=0}^{q-d} b_{\alpha} b_{\alpha+d}, \quad d = 0, 1, \dots, q, \quad (26)$$

and where we define  $b_0 \equiv -1$ .

In order to find the  $M$ -transform of the inverse matrix,  $\mathbf{A}^{(3)} = (\mathbf{A}^{(2)})^{-1}$ , one employs a general result, true for any (real symmetric) random matrix  $\mathbf{H}$ , and obtainable through an easy algebra,

$$M_{\mathbf{H}^{-1}}(z) = -M_{\mathbf{H}}(1/z) - 1. \quad (27)$$

Since the quantity  $M_{\mathbf{A}^{(2)}}(z)$  is known for any  $q$  (42), hence is  $M_{\mathbf{A}^{(3)}}(z)$  via (27), but we will not print it explicitly. Let us just give it for  $q = 1$ , in which case (27) and (21) yield

$$M_{\mathbf{A}^{(3)}}(z) = -\frac{1}{\sqrt{1 - \frac{(1-b_1)^2}{a_0^2} z} \sqrt{1 - \frac{(1+b_1)^2}{a_0^2} z}}. \quad (28)$$

### 3. The Auto-Covariance Matrix

Despite being somewhat outside of the main line of thought of this article, an interesting question would be to search for an explicit expression for the auto-covariance matrix  $\mathbf{A}^{(3)}$  from its Fourier transform (25),

$$A^{(3)}(d) = \frac{1}{2\pi} \int_{-\pi}^{\pi} dp e^{-idp} \frac{1}{\kappa_0^{(2)} + 2 \sum_{l=1}^q \kappa_l^{(2)} \cos(lp)}, \quad (29)$$



where we exploited the fact that  $\mathbf{A}^{(3)}$  must be translationally invariant,  $A_{ab}^{(3)} = A^{(3)}(a - b)$ . This computation would shed light on the structure of temporal correlations present in a VAR setting.

This integral is evaluated by the method of residues in a very similar manner to the one shown in appendix 3, and we do this in appendix 4. We discover that the auto-covariance matrix is a sum of  $q$  exponential decays,

$$A^{(3)}(d) = \sum_{\gamma=1}^q C_{\gamma} e^{-|d|/T_{\gamma}}, \quad (30)$$

where  $C_{\gamma}$  are constants, and  $T_{\gamma}$  are the characteristic times (44),  $\gamma = 1, \dots, q$ ; these constituents are given explicitly in (43). This is a well-known fact, nevertheless we wanted to establish it again within our approach.

For example, for  $q = 1$ , the auto-covariance matrix of VAR(1) is one exponential decay,

$$A^{(3)}(d) = \frac{a_0^2}{1 - b_1^2} b_1^{|d|}, \quad (31)$$

where we assumed for simplicity  $0 < b_1 < 1$  (the formula can be easily extended to all values of  $b_1$ ).

#### 4. The Pearson Estimator of the Covariances from Free Random Variables

Having found an expression for the  $M$ -transform of the auto-covariance matrix  $\mathbf{A}^{(3)}$  of a VAR( $q$ ) (27), (42), we may proceed to investigate the equation (14) for the  $M$ -transform  $M \equiv M_{\mathbf{c}}(z)$  of the correlated Wishart random matrix  $\mathbf{c} = (1/T)\mathbf{Y}\mathbf{Y}^T = (1/T)\tilde{\mathbf{Y}}\mathbf{A}^{(3)}\tilde{\mathbf{Y}}^T$  (4). We will do this explicitly only for  $q = 1$ , when (28) leads to a fourth-order (Ferrari) polynomial equation for the unknown  $M$ ,

$$\begin{aligned} & a_0^4 r^2 M^4 + 2a_0^2 r \left( - (1 + b_1^2) z + a_0^2 r \right) M^3 + \\ & + \left( (1 - b_1^2)^2 z^2 - 2a_0^2 r (1 + b_1^2) z + (r^2 - 1) a_0^4 \right) M^2 - 2a_0^4 M - a_0^4 = 0. \end{aligned} \quad (32)$$

This equation has been derived by another method in [6], and our result confirms their equation (A.8), with the change in notation,  $y \rightarrow 1/r$ ,  $x \rightarrow z/r$ ,  $z \rightarrow rM$ .

### C. The VARMA( $q_1, q_2$ ) Process

#### 1. The Definition of VARMA( $q_1, q_2$ )

The two types of processes which we elaborated on above, VAR( $q_1$ ) and VMA( $q_2$ ), can be combined into one stochastic process called VARMA( $q_1, q_2$ ),

$$Y_{ia} - \sum_{\beta=1}^{q_1} b_{\beta} Y_{i,a-\beta} = \sum_{\alpha=0}^{q_2} a_{\alpha} \epsilon_{i,a-\alpha}. \quad (33)$$

Now it is a straightforward and well-known observation (which can be verified by a direct calculation) that the auto-covariance matrix  $\mathbf{A}^{(5)}$  of this process is simply the product (in any order) of the auto-covariance matrices of the VAR and VMA pieces; more precisely,

$$\mathbf{A}^{(5)} = \left( \mathbf{A}^{(4)} \right)^{-1} \mathbf{A}^{(1)}, \quad (34)$$

where  $\mathbf{A}^{(1)}$  corresponds to the generic VMA( $q_2$ ) model (17), while  $\mathbf{A}^{(4)}$  denotes the auto-covariance matrix of VMA( $q_1$ ) with a slightly different modification of the parameters compared to the previously used, namely  $a_0^{(4)} \equiv 1$ ,  $a_{\beta}^{(4)} \equiv -b_{\beta}$ , for  $\beta = 1, \dots, q_1$ . We have thus already made use here of the fact that the auto-covariance matrix of a VAR process is the inverse of the auto-covariance matrix of a certain corresponding VMA process (24), but the new change in parameters necessary in moving from VAR to VMA has effectively  $a_0 = 1$  w.r.t. what we had before (24); it is understandable: this “missing”  $a_0$  is now included in the matrix of the other VMA( $q_2$ ) process.

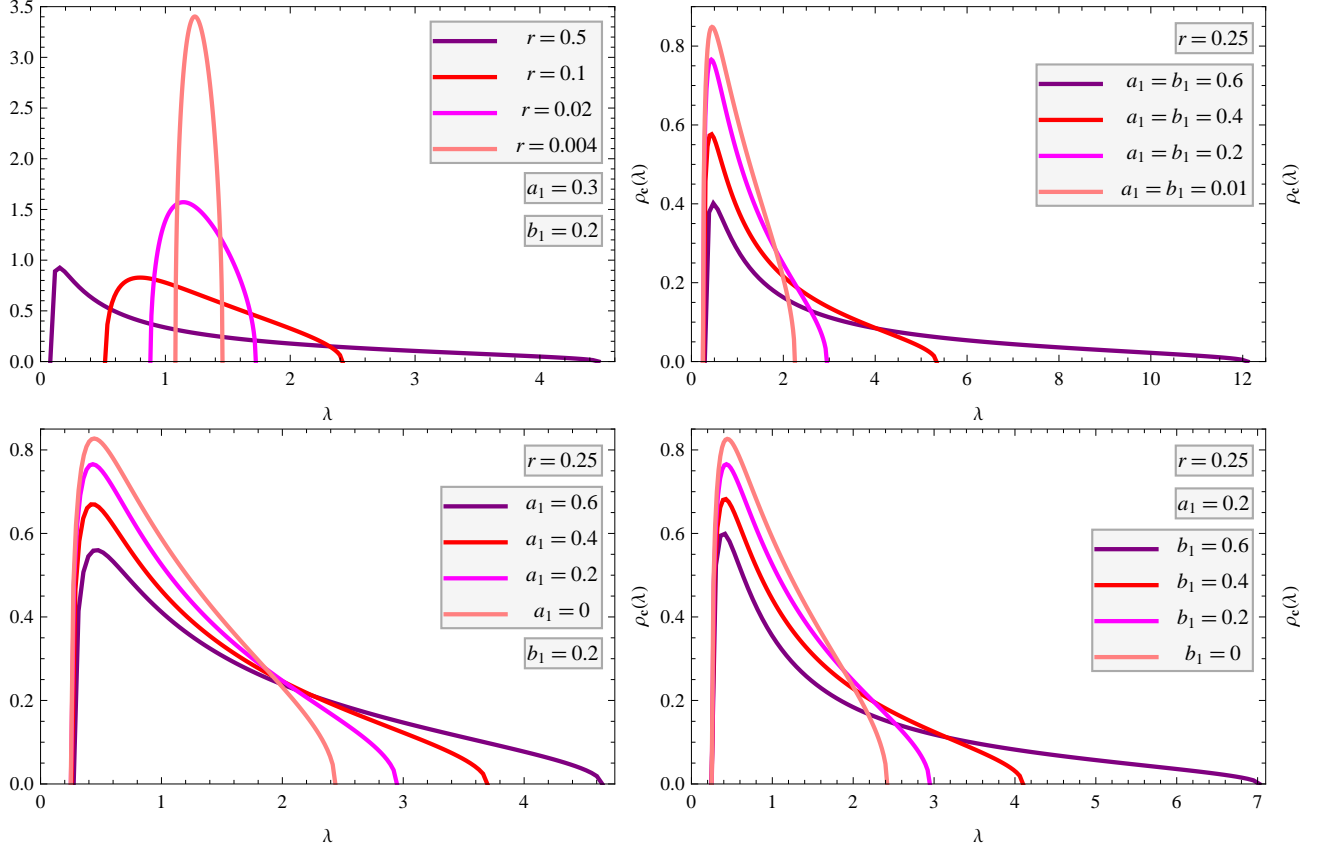


FIG. 1: The mean spectral density  $\rho_{\mathbf{c}}(\lambda)$  of the Pearson estimator  $\mathbf{c}$  of the cross-covariances in the VARMA(1, 1) process computed numerically from the sixth-order polynomial equation (45), for various values of the process' parameters. The scale of these parameters is determined by choosing  $a_0 = 1$  everywhere. Recall that the theoretical formula (45) is valid in the thermodynamical limit (6) of  $N, T \rightarrow \infty$ , with  $r = N/T$  kept finite.

UP LEFT: We set the remaining VARMA parameters to  $a_1 = 0.3$ ,  $b_1 = 0.2$ , while the rectangularity ratio takes the values  $r = 0.5$  (the purple line), 0.1 (red), 0.02 (magenta), 0.004 (pink); each one is 5 times smaller than the preceding one. We observe how the graphs become increasingly peaked (narrower and taller) around  $\lambda = 1$  as  $r$  decreases, which reflects the movement of the estimator  $\mathbf{c}$  toward its underlying value  $\mathbf{C} = \mathbf{1}_N$ .

UP RIGHT: We fix  $r = 0.25$  and consider the two VARMA parameters equal to each other, with the values  $a_1 = b_1 = 0.6$  (purple), 0.4 (red), 0.2 (magenta), 0.01 (pink).

DOWN LEFT: We hold  $r = 0.25$  and  $b_1 = 0.2$ , and modify  $a_1 = 0.6$  (purple), 0.4 (red), 0.2 (magenta), 0.0 (pink); for this last value, the VARMA(1, 1) model reduces to VAR(1).

DOWN RIGHT: Similarly, but this time we assign  $r = 0.25$  and  $a_1 = 0.2$ , while changing  $b_1 = 0.6$  (purple), 0.4 (red), 0.2 (magenta), 0.0 (pink); this last value corresponds to VMA(1).

## 2. The Fourier Transform and the $M$ -Transform of the Auto-Covariance Matrix

The Fourier transform of the auto-covariance matrix  $\mathbf{A}^{(5)}$  of VARMA( $q_1, q_2$ ) (34) is simply the product of the respective Fourier transforms (19) and (25),

$$\widehat{A}^{(5)}(p) = \frac{\kappa_0^{(1)} + 2 \sum_{d_2=1}^{q_2} \kappa_{d_2}^{(1)} \cos(d_2 p)}{\kappa_0^{(4)} + 2 \sum_{d_1=1}^{q_1} \kappa_{d_1}^{(4)} \cos(d_1 p)}, \quad (35)$$

where

$$\kappa_{d_1}^{(4)} = \sum_{\alpha_1=0}^{q_1-d_1} b_{\alpha_1} b_{\alpha_1+d_1}, \quad \kappa_{d_2}^{(1)} = \sum_{\alpha_2=0}^{q_2-d_2} a_{\alpha_2} a_{\alpha_2+d_2}, \quad d_1 = 0, 1, \dots, q_1, \quad d_2 = 0, 1, \dots, q_2, \quad (36)$$

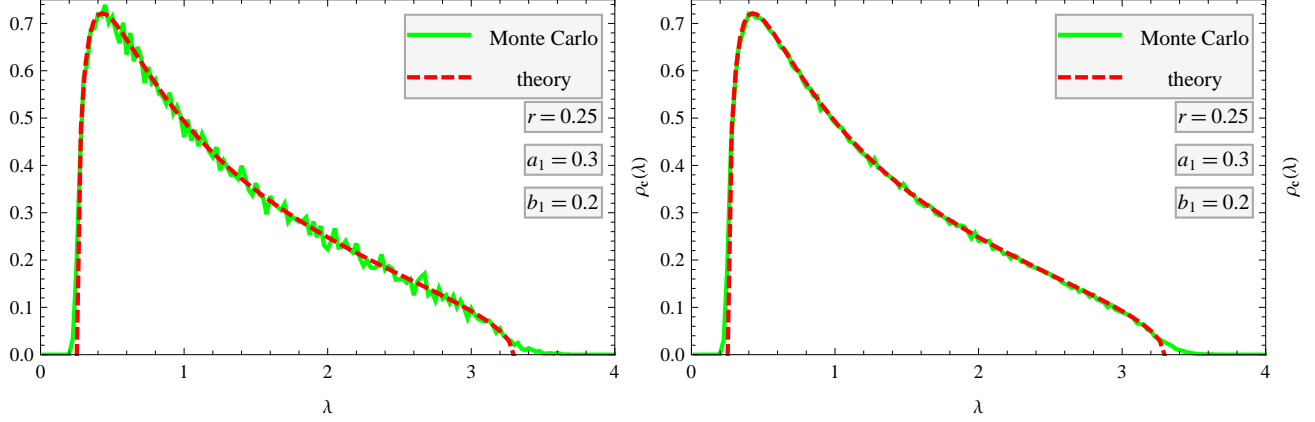


FIG. 2: Monte Carlo simulations of the mean spectral density  $\rho_c(\lambda)$  (the green plots) compared to the theoretical result obtained numerically from the sixth-order equation (45) (the dashed red lines). The conformity is nearly perfect. We generate the matrices  $\mathbf{Y}$  of sizes  $N = 50$ ,  $T = 200$  (i.e.,  $r = 0.25$ ) from the VARMA(1,1) process with the parameters  $a_0 = 1$ ,  $a_1 = 0.3$ ,  $b_1 = 0.2$ . The Monte Carlo simulation is repeated 1,000 (LEFT) or 10,000 (RIGHT) times; in this latter case, a significant improvement in the quality of the agreement is seen. One notices finite-size effects at the edges of the spectrum (“leaking out” of eigenvalues): in the numerical simulations,  $N$  and  $T$  are obviously finite, while equation (45) is legitimate in the thermodynamical limit (6) only, hence the small discrepancies; by enlarging the chosen dimensions  $50 \times 200$  one would diminish this fallout.

where we recall  $b_0 = -1$ . For instance, for VARMA(1,1) (it is described by three constants,  $a_0$ ,  $a_1$ ,  $b_1$ ), one explicitly has

$$\widehat{A^{(5)}}(p) = \frac{a_0^2 + a_1^2 + 2a_0a_1 \cos p}{1 + b_1^2 - 2b_1 \cos p}. \quad (37)$$

The  $M$ -transform of  $\mathbf{A}^{(5)}$  can consequently be derived from the general formula (20). We will evaluate here the pertinent integral only for the simplest VARMA(1,1) process, even though an arbitrary case may be handled by the technique of residues,

$$M_{\mathbf{A}^{(5)}}(z) = \frac{1}{a_0a_1 + b_1z} \left( -a_0a_1 + \frac{z(a_0a_1 + (a_0^2 + a_1^2)b_1 + a_0a_1b_1^2)}{\sqrt{(1-b_1)^2z - (a_0+a_1)^2} \sqrt{(1+b_1)^2z - (a_0-a_1)^2}} \right). \quad (38)$$

### 3. The Auto-Covariance Matrix

One might again attempt to track the structure of temporal covariances in a VARMA process. This can be done either by the inverse Fourier transform of (35), or through a direct computation based on the recurrence relation (33) (importantly, adhering to the assumption that it stretches to the past infinity). Let us print the result just for VARMA(1,1),

$$A^{(5)}(d) = -\frac{a_0a_1}{b_1} \delta_{d,0} + \frac{(a_1 + a_0b_1)(a_0 + a_1b_1)}{b_1(1-b_1^2)} b_1^{|d|}, \quad (39)$$

where for simplicity  $0 < b_1 < 1$ . This is an exponential decay, with the characteristic time of the VAR piece, with an additional term on the diagonal.

### 4. The Pearson Estimator of the Covariances from Free Random Variables

Expression (38), along with the fundamental FRV formula (14), allow us to write the equation satisfied by the  $M$ -transform  $M \equiv M_{\mathbf{c}}(z)$  of the Pearson estimator  $\mathbf{c} = (1/T)\mathbf{Y}\mathbf{Y}^T = (1/T)\widehat{\mathbf{Y}}\mathbf{A}^{(5)}\widehat{\mathbf{Y}}^T$  (4) of the cross-covariances in the

VARMA(1, 1) process; it happens to be polynomial of order six, and we print it (45) in appendix 5. It may be solved numerically, a proper solution chosen (the one which leads to a sensible density: real, positive–definite, normalized to unity), and finally, the mean spectral density  $\rho_c(\lambda)$  derived from (9). We show the shapes of this density for a variety of the values of the parameters  $r, a_0, a_1, b_1$  in fig. 1. Moreover, in order to test the result (45), and more broadly, to further establish our FRV framework in the guise of formula (14), the theoretical form of the density is compared to Monte Carlo simulations in fig. 2; they remain in excellent concord. These are the main findings of this article.

#### IV. CONCLUSIONS

In this paper we attempted to advertise the power and flexibility of the Free Random Variables calculus for multivariate stochastic processes of the VARMA type. The FRV calculus is ideally suited for multidimensional time series problems, provided the dimensions of the underlying matrices are large. The operational procedures are simple, algebraic and transparent. The structure of the final formula which relates the moments’ generating function of the population covariance and the sample covariance allows one to easily derive eigenvalue density of the sample covariance. We in detail illustrated how this procedure works for VARMA(1, 1), confronted the theoretical prediction with numerical data obtained by Monte Carlo simulations of the VARMA process and observed a perfect agreement.

The FRV calculus is not restricted to Gaussian variables. It also works for non–Gaussian processes, including those with heavy–tailed increments belonging to the Lévy basin of attraction, where the moments do not exist. Since the majority of data collected nowadays is naturally stored in the form of huge matrices, we believe that the FRV technique is the most natural candidate for the matrix–valued “probability calculus” that can provide efficient algorithms for cleaning (de–noising) large sets of data and unraveling essential but hidden correlations.

#### Acknowledgments

This work has been supported by the Polish Ministry of Science Grant No. N N202 229137 (2009–2012). AJ acknowledges the support of Clico Ltd.

#### Appendices

##### 1. The Auto–Covariance Matrix for VMA( $q$ )

In this appendix, we sketch a proof of the formula (17) for the auto–covariance matrix of the VMA( $q$ ) process. As mentioned, since the random variables are centered Gaussian, this matrix alone suffices to completely capture all their properties. We set  $i = j$ ; the dependence on this index may be dropped as there are no correlations here. We use the definition (15) of VMA( $q$ ), as well as the auto–covariance structure of the white noise,  $\langle \epsilon_{ia} \epsilon_{jb} \rangle = \delta_{ij} \delta_{ab}$ . This leads to

$$A_{ab}^{(1)} = \langle Y_{ia} Y_{ib} \rangle = \sum_{\alpha=0}^q \sum_{\beta=0}^q a_{\alpha} a_{\beta} \langle \epsilon_{i,a-\alpha} \epsilon_{i,b-\beta} \rangle = \sum_{\alpha=0}^q \sum_{\beta=0}^q a_{\alpha} a_{\beta} \delta_{a-\alpha, b-\beta} = \dots$$

The double sum is symmetrized, the index  $\beta$  replaced by  $d \equiv \beta - \alpha$ ,

$$\dots = \frac{1}{2} \sum_{\alpha=0}^q \sum_{d=-\alpha}^{q-\alpha} a_{\alpha} a_{\alpha+d} (\delta_{b, a+d} + \delta_{b, a-d}) = \dots,$$

and the order of the sums interchanged (an elegant method for this is explained in [20]),

$$\dots = \frac{1}{2} \sum_{d=-q}^q \left( \sum_{\alpha=\max(0, -d)}^{q-\min(0, d)} a_{\alpha} a_{\alpha+d} \right) (\delta_{b, a+d} + \delta_{b, a-d}),$$

which, upon splitting the sum over  $d$  into three pieces (from  $-q$  to  $-1$ ,  $d = 0$ , and from 1 to  $q$ ), is quickly seen to coincide with (17).

## 2. A List of the Various Auto-Covariance Matrices Used

For the reader's convenience, let us collect in this appendix the five auto-covariance matrices which are defined throughout this paper:

- By  $\mathbf{A}^{(1)}$  we denote the auto-covariance matrix of the VMA( $q$ ) process with the generic constants  $a_\alpha$ , with  $\alpha = 0, 1, \dots, q$ , as defined in (15).
- By  $\mathbf{A}^{(2)}$  we denote the auto-covariance matrix of the VMA( $q$ ) process with the constants  $a_0^{(2)} \equiv 1/a_0$ ,  $a_\beta^{(2)} \equiv -b_\beta/a_0$ , where  $\beta = 1, \dots, q$ .
- By  $\mathbf{A}^{(3)}$  we denote the auto-covariance matrix of the VAR( $q$ ) process with the generic constants  $a_0, b_\beta$ , with  $\beta = 1, \dots, q$ , as defined in (23). There holds  $\mathbf{A}^{(3)} = (\mathbf{A}^{(2)})^{-1}$  (24).
- By  $\mathbf{A}^{(4)}$  we denote the auto-covariance matrix of the VMA( $q_1$ ) process with the constants  $a_0^{(4)} \equiv 1$ ,  $a_\beta^{(4)} \equiv -b_\beta$ , where  $\beta = 1, \dots, q_1$ .
- By  $\mathbf{A}^{(5)}$  we denote the auto-covariance matrix of the VARMA( $q_1, q_2$ ) process with the generic constants  $b_\beta, a_\alpha$ , with  $\beta = 1, \dots, q_1$  and  $\alpha = 0, 1, \dots, q_2$ , according to the definition (33). There is  $\mathbf{A}^{(5)} = (\mathbf{A}^{(4)})^{-1}\mathbf{A}^{(1)}$  (34), where in the latter piece  $q = q_2$ .

## 3. The $M$ -Transform of the Auto-Covariance Matrix for VMA( $q$ )

We will derive here the  $M$ -transform (20) of the auto-covariance matrix  $\mathbf{A}^{(1)}$  of an arbitrary VMA( $q$ ) process, using the expression for its Fourier transform (19). It is a little simpler to consider the Green's function,

$$G_{\mathbf{A}^{(1)}}(z) = \frac{1 + M_{\mathbf{A}^{(1)}}(z)}{z} = \frac{1}{\pi} \int_0^\pi dp \frac{1}{z - \widehat{A}^{(1)}(p)}, \quad (40)$$

where the integration range has been halved due to the evenness of the integrand.

This integral is performed with help of the change of variables  $y \equiv 2 \cos p$ . The measure, when  $p \in [0, \pi]$ , reads  $dp = -dy/\sqrt{4-y^2}$ . A basic observation is that the denominator of the integrand is a linear combination of  $\cos(dp)$ , for  $d = 1, \dots, q$ , and each such a cosine can be cast as a polynomial of order  $d$  in  $y$  through the de Moivre formula. Hence, the denominator is a polynomial of order  $q$  in  $y$ ,

$$\widehat{A}^{(1)}(p) - z = \kappa_0^{(1)} - z + 2 \sum_{d=1}^q \kappa_d^{(1)} \cos(dp) = \psi \prod_{\beta=1}^q (y - y_\beta), \quad (41)$$

where the  $y_\beta$ 's are the  $q$  roots (which we assume to be single), and  $\psi$  is the coefficient at  $y^q$ . Using the method of residues, one readily finds

$$G_{\mathbf{A}^{(1)}}(z) = -\frac{1}{\pi} \frac{1}{\psi} \int_{-2}^2 dy \frac{1}{\sqrt{4-y^2}} \frac{1}{\prod_{\beta=1}^q (y - y_\beta)} = \frac{1}{\psi} \sum_{\gamma=1}^q \frac{1}{\prod_{\substack{\beta=1 \\ \beta \neq \gamma}}^q (y_\gamma - y_\beta)} \frac{1}{\sqrt{y_\gamma - 2} \sqrt{y_\gamma + 2}}, \quad (42)$$

where the two square roots on the r.h.s. are principal. This is an explicit formula for the Green's function of  $\mathbf{A}^{(1)}$ , provided one has factorized the order- $q$  polynomial (41).

## 4. The Auto-Covariance Matrix for VAR( $q$ )

Let us argue now that the Fourier transform (25) leads to the auto-covariance matrix of VAR( $q$ ) (29) of the form of a sum of exponential decays (30), and let us give precise expressions for the constants  $C_\gamma$  and the characteristic times  $T_\gamma$ ,  $\gamma = 1, \dots, q$ .

We proceed by the technique of residues, analogously to appendix 3, however this time with aid of another variable,  $x \equiv e^{-ip}$ , related to the previously used through  $y = 2 \cos p = x + 1/x$ . The integration measure is  $dp = idx/x$ , and the integration path is counterclockwise around the centered unit circle. The denominator of the integrand is a polynomial of order  $q$  in the variable  $y$ , having thus some  $q$  roots  $\tilde{y}_\beta$ ,  $\beta = 1, \dots, q$ . Therefore, there are  $2q$  corresponding solutions for the variable  $x$ , with a half of them inside the integration path and a half outside; let  $\tilde{x}_\beta$  be the solutions to  $x + 1/x = \tilde{y}_\beta$  with the absolute values less than 1. Only them contribute to the integral, and their residues straightforwardly give

$$A^{(3)}(d) = \frac{1}{\psi} \sum_{\gamma=1}^q \frac{(\tilde{x}_\gamma)^{|d|+q-1}}{\prod_{\substack{\beta=1 \\ \beta \neq \gamma}}^q (\tilde{x}_\gamma - \tilde{x}_\beta) \prod_{\beta=1}^q \left( \tilde{x}_\gamma - \frac{1}{\tilde{x}_\beta} \right)}. \quad (43)$$

This is indeed  $q$  exponents  $(\tilde{x}_\gamma)^{|d|}$ ,  $\gamma = 1, \dots, q$ . Remark that the solutions may be complex, hence this is really  $q$  different exponential decays  $\exp(-|d|/T_\gamma)$ , with the characteristic times

$$T_\gamma \equiv -\frac{1}{\log |\tilde{x}_\gamma|} \quad (44)$$

(these times are positive as the roots have the absolute values less than 1), possibly modulated by sinusoidal oscillations when a root has an imaginary part.

For example, for  $q = 1$  there is one exponential decay (31), while for  $q = 2$ , one obtains either two exponential decays (the two roots are real and different), or one exponential decay modulated by oscillations (the two roots are complex and mutually conjugate), *etc.*

## 5. The Equation for the $M$ -Transform of the Pearson Estimator of the Covariances for VARMA(1,1)

The sixth-order polynomial equation obeyed by  $M \equiv M_c(z)$  in the case of VARMA(1,1) reads,

$$\begin{aligned} & r^4 a_0^2 a_1^2 (a_0^2 - a_1^2)^2 M^6 + \\ & + 2r^3 a_0 a_1 \left( \left( (a_0^4 - 6a_0^2 a_1^2 + a_1^4) b_1 - a_0 a_1 (a_0^2 + a_1^2) (b_1^2 + 1) \right) z + (1 + 2r) a_0 a_1 (a_0^2 - a_1^2)^2 \right) M^5 + \\ & + r^2 \left( \left( (a_0^4 - 20a_0^2 a_1^2 + a_1^4) b_1^2 - 4a_0 a_1 (a_0^2 + a_1^2) b_1 (b_1^2 + 1) + a_0^2 a_1^2 (b_1^4 + 1) \right) z^2 + \right. \\ & + 2a_0 a_1 \left( ((1 + 3r) (a_0^4 + a_1^4) - 2(5 + 9r) a_0^2 a_1^2) b_1 - (2 + 3r) a_0 a_1 (a_0^2 + a_1^2) (b_1^2 + 1) \right) z + \\ & \left. + (1 + 8r + 6r^2) a_0^2 a_1^2 (a_0^2 - a_1^2)^2 \right) M^4 + \\ & + 2r \left( b_1 \left( -6a_0 a_1 b_1^2 - (a_0^2 + a_1^2) b_1 (b_1^2 + 1) + a_0 a_1 (b_1^4 + 1) \right) z^3 + \right. \\ & + \left( (-10(1 + 2r) a_0^2 a_1^2 + r(a_0^4 + a_1^4)) b_1^2 - 2(1 + 2r) a_0 a_1 (a_0^2 + a_1^2) b_1 (b_1^2 + 1) + (1 + r) a_0^2 a_1^2 (b_1^4 + 1) \right) z^2 + \\ & \left. + a_0 a_1 \left( (3r(1 + r) (a_0^4 + a_1^4) - 2(2 + 15r + 9r^2) a_0^2 a_1^2) b_1 - (1 + 6r + 3r^2) a_0 a_1 (a_0^2 + a_1^2) (b_1^2 + 1) \right) z + \right. \end{aligned}$$

$$\begin{aligned}
& + 2r (1 + 3r + r^2) a_0^2 a_1^2 (a_0^2 - a_1^2)^2 \Big) M^3 + \\
& + \left( b_1^2 (1 - b_1^2)^2 z^4 + 2b_1 \left( -2(1 + 3r) a_0 a_1 b_1^2 - r (a_0^2 + a_1^2) b_1 (b_1^2 + 1) + (1 + r) a_0 a_1 (b_1^4 + 1) \right) z^3 + \right. \\
& \quad + \left( -((1 - r^2) (a_0^4 + a_1^4) + 2(3 + 20r + 10r^2) a_0^2 a_1^2) b_1^2 - \right. \\
& \quad \left. \left. - 2(1 + 4r + 2r^2) a_0 a_1 (a_0^2 + a_1^2) b_1 (b_1^2 + 1) + r(4 + r) a_0^2 a_1^2 (b_1^4 + 1) \right) z^2 + \right. \\
& \quad \left. + 2ra_0 a_1 \left( r(3 + r) (a_0^4 + a_1^4) - 6(2 + 5r + r^2) a_0^2 a_1^2 b_1 - (3 + 6r + r^2) a_0 a_1 (a_0^2 + a_1^2) (b_1^2 + 1) \right) z + \right. \\
& \quad \left. + r^2 (6 + 8r + r^2) a_0^2 a_1^2 (a_0^2 - a_1^2)^2 \Big) M^2 + \\
& \quad + 2 \left( a_0 a_1 b_1 (1 - b_1^2)^2 z^3 + \right. \\
& \quad \left. + \left( - (a_0^4 + a_1^4 + 2(3 + 5r) a_0^2 a_1^2) b_1^2 - 2(1 + r) a_0 a_1 (a_0^2 + a_1^2) b_1 (b_1^2 + 1) + r a_0^2 a_1^2 (b_1^4 + 1) \right) z^2 + \right. \\
& \quad \left. + r a_0 a_1 \left( r (a_0^4 + a_1^4) - 2(6 + 5r) a_0^2 a_1^2 b_1 - (3 + 2r) a_0 a_1 (a_0^2 + a_1^2) (b_1^2 + 1) \right) z + \right. \\
& \quad \left. + r^2 (2 + r) a_0^2 a_1^2 (a_0^2 - a_1^2)^2 \Big) M - \\
& \quad - b_1 \left( (a_0^4 + 6a_0^2 a_1^2 + a_1^4) b_1 + 2a_0 a_1 (a_0^2 + a_1^2) (b_1^2 + 1) \right) z^2 - \\
& \quad - 2ra_0^2 a_1^2 \left( 4a_0 a_1 b_1 + (a_0^2 + a_1^2) (b_1^2 + 1) \right) z + r^2 a_0^2 a_1^2 (a_0^2 - a_1^2)^2 = 0. \tag{45}
\end{aligned}$$

This equation in a *Mathematica* file can be obtained from the authors upon request.

- 
- [1] Sims C. A., *Macroeconomics and reality*, *Econometrica* **48** (1980) 1.  
[2] Smets F., Wouters R., *An Estimated Stochastic Dynamic General Equilibrium Model of the Euro Area*, European Central Bank Working Paper Series, working paper **171**, August 2002 [<http://www.ecb.int>].  
[3] Box G., Jenkins G. M., Reinsel G., *Time Series Analysis: Forecasting and Control*, 3rd ed., Prentice-Hall, Englewood Cliffs, NJ, 1994.

- [4] Laloux L., Cizeau P., Bouchaud J.-P., Potters M., *Noise Dressing of Financial Correlation Matrices*, Phys. Rev. Lett. **83** (1999) 1467 [[arXiv:cond-mat/9810255](#)].
- [5] Plerou V., Gopikrishnan P., Rosenow B., Amaral L. A. N., Stanley H. E., *Universal and non-universal properties of cross-correlations in financial time series*, Phys. Rev. Lett. **83** (1999) 1471 [[arXiv:cond-mat/9902283](#)].
- [6] Jin B., Wang C., Miao B., Huang M.-N. L., *Limiting spectral distribution of large-dimensional sample covariance matrices generated by VARMA*, Journal of Multivariate Analysis **100** (2009) 2112.
- [7] Bai Z. D., Silverstein J. W., *Spectral Analysis of Large Dimensional Random Matrices*, Science Press, Beijing, 2006.
- [8] Voiculescu D. V., Dykema K. J., Nica A., *Free Random Variables*, CRM Monograph Series, Vol. 1, Am. Math. Soc., Providence, 1992.
- [9] Speicher R., *Multiplicative functions on the lattice of non-crossing partitions and free convolution*, Math. Ann. **298** (1994) 611.
- [10] Burda Z., Jarosz A., Jurkiewicz J., Nowak M. A., Papp G., Zahed I., *Applying Free Random Variables to Random Matrix Analysis of Financial Data*, submitted to Quantitative Finance.
- [11] Wishart J., *The Generalized Product Moment Distribution in Samples from a Normal Multivariate Population*, Biometrika **A 20** (1928) 32.
- [12] Zee A., *Law of addition in random matrix theory*, Nucl. Phys. **B 474** (1996) 726 [[arXiv:cond-mat/9602146](#)].
- [13] Janik R. A., Nowak M. A., Papp G., Zahed I., *Various Shades of Blue's Functions*, Acta Phys. Polon. **B 28** (1997) 2949 [[arXiv:hep-th/9710103](#)].
- [14] Burda Z., Görlich A., Jarosz A., Jurkiewicz J., *Signal and Noise in Correlation Matrix*, Physica **A 343** (2004) 295 [[arXiv:cond-mat/0305627](#)].
- [15] Burda Z., Jurkiewicz J., *Signal and Noise in Financial Correlation Matrices*, Physica **A 344** (2004) 67 [[arXiv:cond-mat/0312496](#)].
- [16] Burda Z., Jurkiewicz J., Waclaw B., *Spectral Moments of Correlated Wishart Matrices*, Phys. Rev. **E 71** (2005) 026111 [[arXiv:cond-mat/0405263](#)].
- [17] Burda Z., Görlich A., Jurkiewicz J., Waclaw B., *Correlated Wishart Matrices and Critical Horizons*, Eur. Phys. J. **B 49** (2006) 319 [[arXiv:cond-mat/0508341](#)].
- [18] Burda Z., Jurkiewicz J., Waclaw B., *Eigenvalue density of empirical covariance matrix for correlated samples*, Acta Phys. Pol. **B 36** (2005) 2641 [[arXiv:cond-mat/0508451](#)].
- [19] Lütkepohl H., *New Introduction to Multiple Time Series Analysis*, Springer Verlag, Berlin, 2005.
- [20] Graham R., Knuth D., Patashnik O., *Concrete Mathematics: A Foundation for Computer Science*, Addison-Wesley, 1994.