Journal of Statistical Mechanics:
Theory and Experiment

PAPER

On the equivalence between the kinetic Ising model and discrete autoregressive processes

To cite this article: Carlo Campajola *et al J. Stat. Mech.* (2021) 033412

Journal of Statistical Mechanics: Theory and Experiment
An IOP and SISSA journal

PAPER: Interdisciplinary statistical mechanics

On the equivalence between the kinetic Ising model and discrete autoregressive processes

Carlo Campajola[1,2,3,*], **Fabrizio Lillo**[1,4], **Piero Mazzarisi**[1] **and Daniele Tantari**[4]

[1] Scuola Normale Superiore, Piazza dei Cavalieri 7, 56126 Pisa, Italy
[2] UZH Blockchain Center, University of Zürich, Rämistrasse 71, 8006 Zürich, Switzerland
[3] URPP Social Networks, University of Zürich, Andreasstrasse 15, 8050 Zürich, Switzerland
[4] Department of Mathematics, University of Bologna, Piazza di Porta San Donato 5, 40126 Bologna, Italy
E-mail: carlo.campajola@uzh.ch

Received 27 November 2020
Accepted for publication 8 February 2021
Published 25 March 2021

Online at stacks.iop.org/JSTAT/2021/033412
https://doi.org/10.1088/1742-5468/abe946

Abstract. Binary random variables are the building blocks used to describe a large variety of systems, from magnetic spins to financial time series and neuron activity. In statistical physics the kinetic Ising model has been introduced to describe the dynamics of the magnetic moments of a spin lattice, while in time series analysis discrete autoregressive processes have been designed to capture the multivariate dependence structure across binary time series. In this article we provide a rigorous proof of the equivalence between the two models in the range of a unique and invertible map unambiguously linking one model parameters set to the other. Our result finds further justification acknowledging that both models provide maximum entropy distributions of binary time series with given means, auto-correlations, and lagged cross-correlations of order one. We further show that the equivalence between the two models permits to exploit the inference methods originally developed for one model in the inference of the other.

*Author to whom any correspondence should be addressed.

1742-5468/21/033412+13$33.00

Keywords: kinetic Ising models, exact results, models of financial markets, stochastic processes

Contents

1. Introduction

The dynamics of a large variety of systems, from physics to economics and finance, can be represented as time series of binary variables. The most telling examples are the *spin systems* in statistical physics, where magnetic moments of the particles in a lattice are described as two-state variables, or *binary time series* in quantitative finance, capturing for instance the occurrence of extreme events of prices [1, 2], or buy and sell orders in the order book of financial markets [3]. Different models have been introduced to capture the multivariate interaction structure of such binary systems, in particular the *kinetic Ising model* (KIM) [4] in physics and the *discrete autoregressive processes* [5, 6] in time series analysis, together with the (Markovian) multivariate generalization recently introduced by [7], namely the *vector discrete autoregressive process* VDAR(1). In this paper we prove analytically that, under some condition, the KIM is equivalent to the VDAR(1) model. Furthermore it is well known that the Ising model, in both static [8] and kinetic [9] version, is a maximum entropy model, given mean magnetizations and pairwise correlations (at lag one in the kinetic case), see also [10–12], and, among other aspects, maximum entropy arguments can be used to define without ambiguity the temperature in such nonequilibrium spin systems [13]. Here, by exploiting the equivalence between the two models, we prove also that the Markov chain associated with the VDAR(1) can be interpreted as the maximum entropy distribution of binary random variables with given *means*, *auto-correlations*, and *lagged cross-correlations* (of order one). Thus, the KIM and the VDAR(1) should be preferable to other models in absence of prior information on other metrics, following the principle of maximum entropy.

The KIM [4, 14–16], was originally proposed as the out of equilibrium version of the classical Ising spin glass [17–19] to describe a Markovian dynamics of *spins* σ_t^i, i.e. binary random variables taking values -1 and 1, interacting with each other according to some generic matrix of couplings. The KIM has found countless applications in many contexts such as neuroscience [20, 21], computational biology [22], machine learning [23–25], and economics and finance [26–28].

In mathematical terms, the KIM is a logistic regression model specified by the following transition probability for a set of N spins $\boldsymbol{\sigma}_t \equiv \{\sigma_t^i\}_{i=1,\dots,N}$,

$$p_{\mathrm{KIM}}(\boldsymbol{\sigma}_t|\boldsymbol{\sigma}_{t-1}, \boldsymbol{J}, \boldsymbol{h}) = Z_{t-1}^{-1} \exp\left(\sum_{i,j=1}^{N} \sigma_t^i J_{ij}\sigma_{t-1}^j + \sum_{i=1}^{N} \sigma_t^i h_i\right) \tag{1}$$

where $\boldsymbol{J} \equiv \{J_{ij}\}_{i,j=1,\dots,N}$ is a matrix of real-valued parameters giving the multivariate auto-regressive structure of the model or, equivalently, representing the couplings between spins, $\boldsymbol{h} \equiv \{h_i\}_{i=1,\dots,N}$ is a set of variable-specific parameters representing the external magnetic fields associated with each spin, and Z_{t-1} is the partition function $Z_{t-1} = \prod_{i=1}^{N} 2\cosh\left(\sum_j J_{ij}\sigma_{t-1}^j + h_i\right)$ guaranteeing that the probability distribution is properly normalized.

The KIM of equation (1) is a *maximum entropy* [29, 30] model for a set of binary random variables which display on average given *means*, and both *auto-* and (lagged) *cross-correlations*. For the sake of clarity and in preparation to the section below, let us move from the spin variables $\sigma_t^i \in \{-1, 1\}$ to the binary variables $X_t^i \in \{0, 1\}$.

Given a set of N binary variables $\boldsymbol{X}_t \equiv \{X_t^i\}_{i=1,\dots,N}^{t=1,\dots,T}$, let us consider the following metrics,

$$2\sum_t X_t^i, \quad \forall i = 1, \dots, N, \tag{2}$$

$$2\sum_t X_t^i X_{t-1}^j + (1 - X_t^i)(1 - X_{t-1}^j), \quad \forall i, j = 1, \dots, N, \tag{3}$$

related (under stationarity conditions) to the mean of the binary random variables and the correlation between them, respectively.

The metric (3) for $i = j$ is known as *stability* [31], which is connected with the sample auto-correlation of a binary sequence, i.e. $\sum_t X_t^i X_{t-1}^i$. Similarly, when $i \neq j$ the metric is related to lagged cross-correlations. The maximum entropy probability distribution of $\boldsymbol{X}_1, \boldsymbol{X}_2, \dots, \boldsymbol{X}_T$, i.e. the one maximizing the entropy $-\sum_{\boldsymbol{X}_1,\dots,\boldsymbol{X}_T} p(\boldsymbol{X}_1, \dots, \boldsymbol{X}_T) \log p(\boldsymbol{X}_1, \dots, \boldsymbol{X}_T)$ while preserving on average some given values for the metrics (2) and (3), has transition probability (by assuming a given initial condition $\boldsymbol{X}_0$ and exploiting the Markov property)

$$p(\boldsymbol{X}_t|\boldsymbol{X}_{t-1}; \boldsymbol{J}, \boldsymbol{h}) = \prod_{i=1}^{N} \frac{\exp\left[2X_t^i\left(h_i + \sum_{j=1}^{N} J_{ij}(2X_{t-1}^j - 1)\right)\right]}{1 + \exp\left[2\left(h_i + \sum_{j=1}^{N} J_{ij}(2X_{t-1}^j - 1)\right)\right]} \tag{4}$$

where $\boldsymbol{J} = \{J_{ij}\}_{i,j=1,\dots,N}$ and $\boldsymbol{h} = \{h_i\}_{i=1,\dots,N}$ are $N^2 + N$ Lagrange multipliers solving the maximum entropy problem [29, 32]. It is trivial to show that the transition probability of the KIM (1) can be stated equivalently as the maximum entropy probability (4) in terms of the binary variables $\boldsymbol{X}_t \forall i, t$, through the relation $X_t^i = \frac{1+\sigma_t^i}{2}$ (with the same parameters $\boldsymbol{J}$ and $\boldsymbol{h}$).

The VDAR(1) model describes the dependence structure of a set of binary random variables which has Markov property and Bernoulli marginal distribution. It has been proposed originally in its univariate version [5] and followed by several extensions such as the Discrete AutoRegressive Moving Average (DARMA) model [6] and recently proposed in its multivariate formulation [7], the VDAR model. Models from this family have seen applications in genetics [33], queueing theory [34], temporal networks [35] and recently in financial systems, as methods to forecast order flows [36] or to identify preferential lending between banks [37].

In terms of the N binary variables $\boldsymbol{X}_t \equiv \{X_t^i\}_{i=1,\dots,N}$ with $X_t^i \in \{0,1\}$ (and initial condition $\boldsymbol{X}_0$), the VDAR(1) process describes the evolution of X_t^i as

$$X_t^i = V_t^i X_{t-1}^{A_t^i} + \left(1 - V_t^i\right) Z_t^i \tag{5}$$

with $V_t^i \sim \mathcal{B}(\nu_i)$ a Bernoulli random variable with parameter $\nu_i \in [0,1]$, $A_t^i \sim \mathcal{M}(\lambda_{i1}, \dots \lambda_{iN})$ a multinomial random variable taking integer value in $\{1, \dots, N\}$, with parameters $\lambda_{i1}, \dots, \lambda_{iN}$ such that $\sum_{j=1}^N \lambda_{ij} = 1$, and $Z_t^i \sim \mathcal{B}(\chi_i)$ with $\chi_i \in [0,1]$. In other words, the VDAR(1) process captures the (multivariate) mechanism of copying from the past: with probability ν_i, X_t^i is copied from the past and, in this case, λ_{ij} is the probability that X_t^i is equal to X_{t-1}^j (including also the past itself with probability λ_{ii}); otherwise, with probability $1-\nu_i$, X_t^i is not copied and is instead sampled according to a Bernoulli marginal with probability χ_i. Hence, the VDAR(1) model describes N binary random variables with both Markov property and some autoregressive dependency structure, similarly to the KIM. The model is formalized by the transition probability

$$p_{\mathrm{VDAR}}(\boldsymbol{X}_t | \boldsymbol{X}_{t-1}\,; \boldsymbol{\pi}) = \prod_{i=1}^N \left[\nu_i \left(\sum_{j=1}^N \lambda_{ij} \delta_{X_t^i, X_{t-1}^j}\right) + (1-\nu_i)(\chi_i)^{X_t^i}(1-\chi_i)^{1-X_t^i}\right] \tag{6}$$

where $\delta_{X_t^i, X_{t-1}^j}$ is the Kronecker delta, $\boldsymbol{\pi} = \{\{\nu_i\}, \{\lambda_{ij}\}, \{\chi_i\}\}_{i,j=1,\dots,N}$.

Notice that the model has $N^2 + N$ parameters, exactly as the KIM. It is thus immediate to ask the question whether a mapping between the two models exists, as well as finding under which conditions the two models can be considered equivalent. In the following, we indicate the KIM model as $\{\{\boldsymbol{X}_t\}, p_{\mathrm{KIM}}, \boldsymbol{\theta}\}$ with set of parameters $\boldsymbol{\theta} \equiv (\boldsymbol{J}, \boldsymbol{h})$, while the VDAR(1) model is summarized as $\{\{\boldsymbol{X}_t\}, p_{\mathrm{VDAR}}, \boldsymbol{\pi}\}$. Calling $\Theta = \mathbb{R}^{N\times N} \times \mathbb{R}^N$ the space of all possible KIM parameters $\boldsymbol{\theta}$ and Π the space of all possible VDAR(1) parameters $\boldsymbol{\pi}$,

Definition 1. The KIM and the VDAR(1) models are said to be equivalent on $(\hat{\Theta}, \hat{\Pi})$ if there exist an unique invertible map $f : \hat{\Pi} \subseteq \Pi \to \hat{\Theta} \subseteq \Theta$ such that

$$p_{\mathrm{KIM}}(\boldsymbol{X}_t | \boldsymbol{X}_{t-1}\,; f(\boldsymbol{\pi})) = p_{\mathrm{VDAR}}(\boldsymbol{X}_t | \boldsymbol{X}_{t-1}\,; \boldsymbol{\pi})$$

for any $\boldsymbol{X}_t$ and $\boldsymbol{X}_{t-1}$.

2. Model equivalence

Before stating the main theorem, let us show that this mapping exists in the trivial cases of $N=1$ and $N=2$. For $N=1$, both J and h are scalar parameters, while the DAR(1) model, namely the univariate version of VDAR(1), has two parameters, i.e. ν and χ ($\lambda=1$ by design, since we can copy only the past value of the single variable), thus it is trivial to prove that

$$h=\frac{1}{4}\,\log\left(\frac{\frac{\chi}{1-\chi}+\nu}{\frac{1}{\chi}-(1-\nu)}\right) \tag{7a}$$

$$J=\frac{1}{4}\,\log\left(1+\frac{\nu}{(1-\nu)^2\chi(1-\chi)}\right). \tag{7b}$$

One can notice that here J is strictly positive as long as $\nu,\chi>0$, while it is $J=0$ if and only if $\nu=0$: this suggests that the VDAR(1) model is indeed a restricted version of the KIM, with the elements of the coupling matrix restricted to positive values. Intuitively, in the KIM $J_{ij}<0$ implies that spin i tends to take the opposite value of the past state of j, whereas the VDAR model describes only positive (or zero) correlations.

When $N=2$, both models have 6 free parameters, three parameters associated with each variable (spin) $i=1,2$, thus one can map the two models by considering three independent configurations of $\boldsymbol{X}_{t-1}$, e.g. $\{X^1_{t-1},X^2_{t-1}\}=\{1,1\},\{0,1\},\{0,0\}$ and one possible realization of X^i_t, e.g. $X^i_t=1$, for both cases $i=1$ and $i=2$. Then, by matching the transition probabilities for the two models, one obtains the following system

$$\begin{pmatrix}1 & 1 & -1\\ 1 & -1 & -1\\ -1 & -1 & -1\end{pmatrix}\begin{pmatrix}J_{i1}\\ J_{i2}\\ h_i\end{pmatrix}=\frac{1}{2}\begin{pmatrix}\log\left(\frac{1}{(1-\nu_i)\chi_i}-1\right)\\ \log\left(\frac{1}{\nu_i(1-\lambda_i)+(1-\nu_i)\chi_i}-1\right)\\ \log\left(\frac{1}{\nu_i+(1-\nu_i)\chi_i}-1\right)\end{pmatrix} \tag{8}$$

$\forall i=1,2$.

Hence, there exists a *unique* mapping $f\colon\Pi\to\Theta$ as long as the linear system of equation (8) admits a solution in the domain of parameters: (i) the solution exists because the matrix in the left-hand side of (8) is invertible, then (ii) the mapping f admits the inverse $f^{-1}\colon\Theta|_{J_{ij}\geqslant 0}\to\Pi$ in the restricted codomain $J_{ij}\geqslant 0\forall i,j$ (as can be verified by simple computations).

Given these premises, we can now move to the main result of this paper, by stating

Theorem 1. *The kinetic Ising model* $\{\{\mathbf{X}_t\},p_{\mathrm{KIM}},\boldsymbol{\theta}\}$ *is equivalent to the VDAR(1) model* $\{\{\mathbf{X}_t\},p_{\mathrm{VDAR}},\boldsymbol{\pi}\}$ *if and only if* $J_{ij}\geqslant 0\ \forall i,j$.

In order to prove the theorem above, let us first prove the existence of a map $f\colon\Pi\to\Theta$ from the VDAR(1) model to the KIM, for any set of parameters $\boldsymbol{\pi}\in\Pi$. In particular,

J. Stat. Mech. (2021) 033412

this map is *unique* because of the linearity of the mapping problem, see below. Second, we prove that the mapping of parameters is *invertible* on its codomain $f(\Pi) \subset \Theta$, corresponding to the set of positive couplings $J_{ij} \geqslant 0$, $\forall i, j$. Thus, the two models are equivalent under such condition.

Let us start by constructing the system of equations generating the mapping for the generic case $N > 2$. Following the same procedure used to obtain equation (8), we find

$$M_n \cdot \begin{pmatrix} J_{i1} \\ J_{i2} \\ J_{i3} \\ \vdots \\ J_{iN} \\ h_i \end{pmatrix} \equiv \begin{pmatrix} 1 & 1 & \dots & 1 & 1 & -1 \\ 1 & 1 & \dots & 1 & -1 & -1 \\ 1 & 1 & \dots & -1 & -1 & -1 \\ 1 & \dots & \dots & \dots & \dots & -1 \\ 1 & -1 & \dots & -1 & -1 & -1 \\ -1 & -1 & \dots & -1 & -1 & -1 \end{pmatrix} \begin{pmatrix} J_{i1} \\ J_{i2} \\ J_{i3} \\ \vdots \\ J_{iN} \\ h_i \end{pmatrix} = \frac{1}{2} \begin{pmatrix} \log\left(\frac{1}{(1-\nu_i)\chi_i} - 1\right) \\ \log\left(\frac{1}{\nu_i \lambda_{iN} + (1-\nu_i)\chi_i} - 1\right) \\ \vdots \\ \vdots \\ \log\left(\frac{1}{\nu_i \sum_{j\geqslant 2} \lambda_{ij} + (1-\nu_i)\chi_i} - 1\right) \\ \log\left(\frac{1}{\nu_i + (1-\nu_i)\chi_i} - 1\right) \end{pmatrix} \quad (9)$$

$\forall i = 1, \dots, N$.

Similarly to the case $N = 2$, the above system is obtained by considering $n \equiv N + 1$ independent configurations for $\mathbf{X}_{t-1}$ and the transition probability associated with $X_t^i = 1$. By matching the transition probabilities (4) and (6) associated with the $N + 1$ independent configurations, one finds the system of equation (9) for variable i. Then, one can repeat the same procedure for all is, thus obtaining N systems of $(N + 1)$ linear equations in $(N + 1)$ unknowns, namely $J_{i1}, J_{i2}, \dots, h_i$, each one characterized by the same matrix M_n in equation (9).

Defining $\Lambda(x) \equiv \frac{e^{2x}}{1+e^{2x}}$, the matching of probabilities associated with the $N + 1$ independent configurations read as

$$\begin{cases} \Lambda\left(h_i + \sum_{j\geqslant 1} J_{ij}\right) = \nu_i + (1-\nu_i)\,\chi_i & \text{if } X_{t-1}^1 = 1, X_{t-1}^2 = 1, \dots, X_{t-1}^N = 1 \\ \Lambda\left(h_i - J_{i1} + \sum_{j>1} J_{ij}\right) = \nu_i \left(\sum_{j=2}^{N} \lambda_{ij}\right) + (1-\nu_i)\,\chi_i & \text{if } X_{t-1}^1 = 0, X_{t-1}^2 = 1, \dots, X_{t-1}^N = 1; \\ \Lambda\left(h_i - \sum_{j\leqslant 2} J_{ij} + \sum_{j>2} J_{ij}\right) = \nu_i \left(\sum_{j=3}^{N} \lambda_{ij}\right) + (1-\nu_i)\,\chi_i & \text{if } X_{t-1}^1 = 0, X_{t-1}^2 = 0, \dots, X_{t-1}^N = 1; \\ \dots & \dots \\ \Lambda\left(h_i - \sum_{j\leqslant n} J_{ij}\right) = (1-\nu_i)\,\chi_i & \text{if } X_{t-1}^1 = 0, X_{t-1}^2 = 0, \dots, X_{t-1}^N = 0, \end{cases} \quad (10)$$

then, equation (9) is obtained by applying $\Lambda^{-1}(y) \equiv \frac{1}{2}\log\left(\frac{1}{y} - 1\right)$ to both sides of equation (10).

Given this result, a unique mapping $f\colon \Pi \to \Theta$ exists as long as there exists the inverse of the matrix M_n in equation (9), i.e. if the determinant of M_n is non-zero. We then start by proving the following

Proposition 1. *Given the determinant of the matrix M_{n-1}, then the determinant of the matrix M_n is*

$$\det(M_n) = (-1)^n 2\,\det(M_{n-1}). \tag{11}$$

Proof of Proposition 1. By means of the minor expansion formula (by using the minors associated with the elements of the first row), the determinant of M_n can be computed as

$$\begin{aligned}\det(M_n) = (+1)1 &\begin{vmatrix} 1 & \dots & 1 & -1 & -1 \\ 1 & \dots & -1 & -1 & -1 \\ \dots & \dots & \dots & \dots & \dots \\ -1 & \dots & -1 & -1 & -1 \\ -1 & \dots & -1 & -1 & -1 \end{vmatrix} + (-1)1 \begin{vmatrix} 1 & \dots & 1 & -1 & -1 \\ 1 & \dots & -1 & -1 & -1 \\ \dots & \dots & \dots & \dots & \dots \\ 1 & \dots & -1 & -1 & -1 \\ -1 & \dots & -1 & -1 & -1 \end{vmatrix} \\ &+ \dots + (-1)^n(1) \begin{vmatrix} 1 & 1 & \dots & 1 & -1 \\ 1 & 1 & \dots & -1 & -1 \\ \dots & \dots & \dots & \dots & \dots \\ 1 & -1 & \dots & -1 & -1 \\ -1 & -1 & \dots & -1 & -1 \end{vmatrix} \\ &+ (-1)^{n+1}(-1) \begin{vmatrix} 1 & 1 & \dots & 1 & -1 \\ 1 & 1 & \dots & -1 & -1 \\ \dots & \dots & \dots & \dots & \dots \\ 1 & -1 & \dots & -1 & -1 \\ -1 & -1 & \dots & -1 & -1 \end{vmatrix}.\end{aligned} \tag{12}$$

In the previous formula, one can notice that the first $n-2$ minors of the sum in the right-hand side are zero, because the last two columns of each $(n-1)\times(n-1)$ matrix are indeed equal (two $(n-1)\times 1$ vectors of -1). Thus, equation (12) is simplified as

$$\det(M_n) = (-1)^n 2 \begin{vmatrix} 1 & 1 & \dots & 1 & -1 \\ 1 & 1 & \dots & -1 & -1 \\ \dots & \dots & \dots & \dots & \dots \\ 1 & -1 & \dots & -1 & -1 \\ -1 & -1 & \dots & -1 & -1 \end{vmatrix} = (-1)^n 2\,\det(M_{n-1}) \tag{13}$$

where we notice that the last two minors of (12) are equal to each other and correspond to the determinant of M_{n-1}. Equation (13) then completes the proof of the proposition.

Thanks to this result, we are now able to prove the existence of the mapping from the VDAR(1) model to the KIM model, expressed by

Proposition 2. *Given $\boldsymbol{\pi} \in \Pi$, there exists a solution of the problem of equation (9) for any $N > 0$ and this solution is unique.*

J. Stat. Mech. (2021) 033412

Proof of Proposition 2. For $N = 1$, the solution can be explicitly computed as showed in equations (7a) and (7b). For $N = 2$ the problem in equation (9) is equivalent to equation (8) and $\det(M_3) = 4$, thus there exists the inverse of the matrix M_3 and the solution is uniquely determined by solving the linear system of equation (8). Because of proposition 1, the determinant of M_n is different from zero, in particular

$$\begin{aligned}\det(M_n) &= (-1)^{\sum_{l=4}^{n} l}\left(2^{n-3}\right)\det(M_3)\\ &= (-1)^{\sum_{l=4}^{n} l}\left(2^{n-3}\right)4,\end{aligned}$$

$\forall n > 3$ (or, equivalently, $\forall N > 2$), thus resulting in the existence of the inverse matrix of M_n. Hence, the solution of the problem in equation (9) can be uniquely determined. This completes the proof of the proposition.

Proof of Theorem 1. Given propositions 1 and 2, we have proved there exists a unique mapping $f\colon \Pi \to \Theta$ from the VDAR(1) model to the KIM. To complete the proof, we are now left with the existence of the *inverse* of the map f in its codomain $f(\Pi) \subset \Theta$. By restricting to such subset of parameters, the two models are equivalent. In particular, we now prove the last claim of theorem 1 which states that the two models are equivalent if and only if $J_{ij} \geqslant 0\ \forall i, j$.

Let us start by proving that, if $\boldsymbol{\pi} \in \Pi$, then $J_{ij} \geqslant 0$. To this end let us go back to equation (10) and notice that, combining the equations by taking the difference between the first and the second, between the second and the third and so on, we obtain the following N relations

$$\begin{cases} \nu_i\left(1 - \sum\limits_{j\geqslant 2}\lambda_{ij}\right) = \Lambda\left(h_i + \sum\limits_{j\geqslant 2} J_{ij} + J_{i1}\right) - \Lambda\left(h_i + \sum\limits_{j\geqslant 2} J_{ij} - J_{i1}\right)\\ \dots\\ \nu_i\lambda_{ik} = \Lambda\left(h_i - \sum\limits_{j<k} J_{ij} + \sum\limits_{j\geqslant k+1} J_{ij} + J_{ik}\right) - \Lambda\left(h_i - \sum\limits_{j<k} J_{ij} + \sum\limits_{j\geqslant k+1} J_{ij} - J_{ik}\right)\\ \dots\\ \nu_i\lambda_{iN} = \Lambda\left(h_i - \sum\limits_{j<N} J_{ij} + J_{iN}\right) - \Lambda\left(h_i - \sum\limits_{j<N} J_{ij} - J_{iN}\right). \end{cases} \tag{14}$$

By definition $\nu_i\lambda_{ij} \geqslant 0$ for any i, j (because it represents a probability), then it is

$$\Lambda(C + J_{ij}) - \Lambda(C - J_{ij}) \geqslant 0.$$

Since $\Lambda(x)$ is a monotonically increasing function of x, the previous inequality is fulfilled if and only if $J_{ij} \geqslant 0 \forall i, j$. Thus, this condition is necessarily true if $\boldsymbol{\pi}$ is in the domain of f.

By following the same steps in the opposite direction it is straightforward to prove the reverse relation, that is $J_{ij} \geqslant 0$ is a sufficient condition to have $f^{-1}(\boldsymbol{\theta}) \in \Pi$. Indeed for

J. Stat. Mech. (2021) 033412

any $J_{ij} \geqslant 0$, the product $\nu_i\lambda_{ij}$ is $0 \leqslant \nu_i\lambda_{ij} \leqslant 1 \forall i, j$ given the system (14) and $\Lambda(x) \in [0, 1]$ for any x. Then, by summing all the equations in system (14), one obtains

$$\nu_i = \Lambda\left(h_i + \sum_j J_{ij}\right) - \Lambda\left(h_i - \sum_j J_{ij}\right)$$

which is also positive and smaller than 1 if $J_{ij} \geqslant 0\ \forall j$. Then, it follows that all the λ_{ij} are $0 \leqslant \lambda_{ij} \leqslant 1\ \forall i, j$. Finally, combining the first and last lines of equation (10), one finds that $0 \leqslant \chi_i \leqslant 1$. This procedure can be repeated for all variables $i = 1, \ldots, N$, thus obtaining the inverse mapping $f^{-1} : \Theta|_{J_{ij}\geqslant 0} \to \Pi$ in the subset of the codomain Θ defined by the condition $J_{ij} \geqslant 0 \forall i, j = 1, \ldots, N$. This concludes the proof.

3. Practical implications in model inference

Having formally demonstrated the equivalence between the KIM and the VDAR opens the door to cross-contamination between the literatures in which they were developed. In this section we show that one can use inference methods of the KIM developed in the statistical literature, namely the mean field (MF) method [38], for improving standard inference methods for discrete autoregressive processes, namely the Yule-Walker (YW) equations.

Specifically, a popular method in time series literature for the inference of autoregressive models is the method of moments, which generates the so-called YW equations matching the empirically measured moments with the ones implied by the model parameters [39]. In the context of the VDAR(1) it can be shown [7] that the YW equations read

$$\mathbb{E}\left(\mathbf{X}_t\right) = \boldsymbol{\phi} + \boldsymbol{\Psi}\mathbb{E}\left(\mathbf{X}_{t-1}\right) \tag{15}$$

$$\mathbb{E}\left(\tilde{\mathbf{X}}_t\tilde{\mathbf{X}}_{t-1}\right) = \boldsymbol{\Psi}\mathbb{E}\left(\tilde{\mathbf{X}}_{t-1}\tilde{\mathbf{X}}_{t-1}\right) \tag{16}$$

where $\tilde{\mathbf{X}}_t = \mathbf{X}_t - \mathbb{E}(\mathbf{X}_t)$, $\boldsymbol{\phi}$ is a N-dimensional vector and $\boldsymbol{\Psi}$ is a $N \times N$ matrix. Solving these linear systems for $\boldsymbol{\phi}$ and $\boldsymbol{\Psi}$ allows to obtain the VDAR parameters as $\nu_i = \sum_j \boldsymbol{\Psi}_{ij}$, $\lambda_{ij} = \boldsymbol{\Psi}_{ij}/\nu_i$ and $\chi_i = \boldsymbol{\phi}_i/(1 - \nu_i)$.

On the other hand, the statistical mechanics literature has developed suitable approximation methods of maximum likelihood estimation of KIM. In particular, we consider the method developed by Mézard and Sakellariou [38], which takes advantage of a MF approximation to infer the parameters of the KIM. The method is exact if the J generating the data has all $J_{ij} \neq 0$ and can be adapted to a sparse version through ℓ_1-regularization or decimation methods [40]. It has been extensively used in applications of the model to real financial and neural data [28, 41].

The equivalence between KIM and VDAR we proved in this paper allows us to use the MF method for doing inference on a VDAR model. To test this idea, we apply MF and YW method on simulated data and compare them both in speed and accuracy. The speed is measured by the time needed to perform the inference as a function of N on a regular commercial laptop, while the accuracy is measured by the bias and root

J. Stat. Mech. (2021) 033412

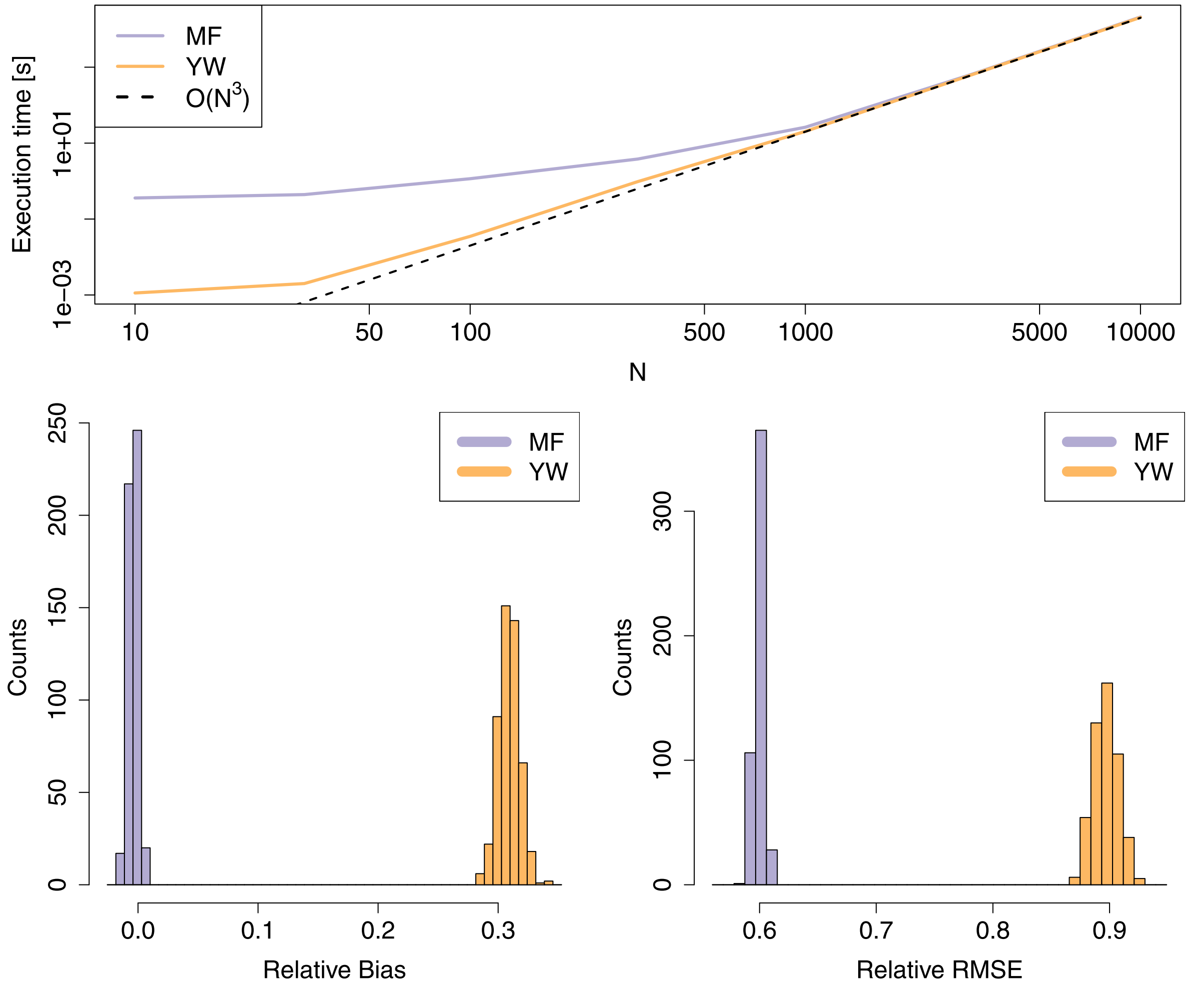

Figure 1. Comparison between the mean field (MF) and Yule-Walker (YW) methods for the inference of the KIM/VDAR(1). (Top) Execution time varying the number of variables N; (bottom left) histogram of the estimator bias relative to the average size of J_{ij} over 500 simulations; (bottom right) histogram of the estimator RMSE relative to the average size of J_{ij} over 500 simulations.

mean squared error (RMSE) of the estimator of J over 500 simulations. We simulate the KIM with uniformly distributed $J_{ij} \sim \mathcal{U}(1/2N, 1/N)$ to keep the model far from the dynamic ferromagnetic transition [15], for $T = 10N$ time steps and N ranging from 10 to 10 000. For the sake of simplicity we consider $h_i = 0\ \forall i$ in our simulations. We show the results of the comparison in figure 1, where it is clear that the YW method is faster on small scale systems but is also less accurate. In particular we see that the method of moments has a positive bias and a relatively large RMSE, whereas the MF method has close to zero bias and a smaller RMSE. The computational effort required in the two methods is comparable as the biggest contribution comes from the inversion of the covariance matrix, typically achieved in $O(N^3)$ operations, hence they present similar execution time for large N. Thus the MF method is a better choice for the

J. Stat. Mech. (2021) 033412

inference of the KIM/VDAR(1), and this result is somewhat expected, since MF is a likelihood-based method whereas the YW equations are not.

This simple analysis shows that the equivalence between the two models can be leveraged to identify inference methods, originally developed for the KIM, that can be used in the inference of the VDAR. As shown above, this can lead to significant improvements in performance when compared to standard VDAR inference methods.

4. Conclusions

In conclusion, the VDAR(1) model is equivalent to the KIM thanks to the existence of a unique mapping for both the binary random variables and the parameters as long as the $\boldsymbol{J}$ parameters of the KIM are positive or zero, as a consequence of the fact that the $\boldsymbol{\nu}$ and $\boldsymbol{\lambda}$ parameters only account for non-negative lagged correlations among random variables. Moreover, since the two models can be interpreted as the maximum entropy distribution of binary random sequences with given means, and both auto- and cross-correlations (only non-negative correlations for the specific case of the VDAR(1) model), both of them represent further the best choice in describing such binary random sequences in absence of prior information on other metrics, according to the principle of maximum entropy.

There are several directions in which future research can go to take advantage of our equivalence theorem. We have already shown that inference methods developed in the statistical physics literature can be used to improve model estimation; another straightforward application of this theorem is that of defining an extension of the VDAR(1) including negative correlations, as the equivalent to the KIM without the restriction on the positive J elements. Finally, it is common in autoregressive models to consider Markov chains of order higher than 1, that is where there are explicit parameters linking the value of X_t to the value of X_{t-k} with $k = 1, \ldots, p$, $p > 1$. The properties of a KIM with higher order interactions have not been studied yet to the best of our knowledge, thus opening another interesting perspective to explore in the context of this cross-contamination between very active research fields.

Acknowledgments

F L acknowledges partial support by the European Integrated Infrastructure for Social Mining and Big Data Analytics (SoBigData++, Grant Agreement #871042). D T acknowledges GNFM-Indam for financial support.

References

[1] Calcagnile L M, Bormetti G, Treccani M, Marmi S and Lillo F 2018 Collective synchronization and high frequency systemic instabilities in financial markets *Quant. Finance* **18** 237–47

[2] Hong Y, Liu Y and Wang S 2009 Granger causality in risk and detection of extreme risk spillover between financial markets *J. Econometrics* **150** 271–87

[3] Bouchaud J-P, Mézard M and Potters M 2002 Statistical properties of stock order books: empirical results and models *Quant. Finance* **2** 251–6

[4] Fredrickson G H and Andersen H C 1984 Kinetic Ising model of the glass transition *Phys. Rev. Lett.* **53** 1244
[5] Jacobs P A and Lewis P A W 1978 Discrete time series generated by mixtures. III. Autoregressive processes (DAR (p)) *Technical Report* Naval Postgraduate School Monterey Calif
[6] Jacobs P A and Lewis P A W 1983 Stationary discrete autoregressive-moving average time series generated by mixtures *J. Time Ser. Anal.* **4** 19–36
[7] Mazzarisi P, Zaoli S, Campajola C and Lillo F 2020 Tail granger causalities and where to find them: extreme risk spillovers vs spurious linkages *J. Econ. Dyn. Control* **121** 104022
[8] Schneidman E, Berry M J, Segev R and Bialek W 2006 Weak pairwise correlations imply strongly correlated network states in a neural population *Nature* **440** 1007–12
[9] Marre O, El Boustani S, Frégnac Y and Destexhe A 2009 Prediction of spatiotemporal patterns of neural activity from pairwise correlations *Phys. Rev. Lett.* **102** 138101
[10] Jaynes E T 1982 On the rationale of maximum-entropy methods *Proc. IEEE* **70** 939–52
[11] Pressé S, Ghosh K, Lee J and Dill K A 2013 Principles of maximum entropy and maximum Caliber in statistical physics *Rev. Mod. Phys.* **85** 1115
[12] Marcaccioli R and Livan G 2020 Correspondence between temporal correlations in time series, inverse problems, and the spherical model *Phys. Rev.* E **102** 012112
[13] Sastre F, Dornic I and Chaté H 2003 Nominal thermodynamic temperature in nonequilibrium kinetic Ising models *Phys. Rev. Lett.* **91** 267205
[14] Bernard D, Gardner E and Zippelius A 1987 An exactly solvable asymmetric neural network model *Europhys. Lett.* **4** 167
[15] Crisanti A and Sompolinsky H 1988 Dynamics of spin systems with randomly asymmetric bonds: Ising spins and Glauber dynamics *Phys. Rev.* A **37** 4865
[16] Sides S W, Rikvold P A and Novotny M A 1998 Kinetic Ising model in an oscillating field: Finite-size scaling at the dynamic phase transition *Phys. Rev. Lett.* **81** 834
[17] Ernst I 1925 Beitrag zur theorie des ferromagnetismus *Z. Phys.* A **31** 253–8
[18] Edwards S F and Anderson P W 1975 Theory of spin glasses *J. Phys. F: Met. Phys.* **5** 965
[19] Scott K and Sherrington D 1978 Infinite-ranged models of spin-glasses *Phys. Rev.* B **17** 4384
[20] Roudi Y and Hertz J 2011 Mean field theory for nonequilibrium network reconstruction *Phys. Rev. Lett.* **106** 048702
[21] Capone C, Filosa C, Gigante G, Ricci-Tersenghi F and Del Giudice P 2015 Inferring synaptic structure in presence of neural interaction time scales *PloS One* **10** e0118412
[22] Imparato A, Pelizzola A and Zamparo M 2007 Ising-like model for protein mechanical unfolding *Phys. Rev. Lett.* **98** 148102
[23] Coolen A C C and Sherrington D 1993 Dynamics of fully connected attractor neural networks near saturation *Phys. Rev. Lett.* **71** 3886
[24] Dunn B and Roudi Y 2013 Learning and inference in a nonequilibrium Ising model with hidden nodes *Phys. Rev.* E **87** 022127
[25] Campajola C, Lillo F and Tantari D 2019 Inference of the kinetic Ising model with heterogeneous missing data *Phys. Rev.* E **99** 062138
[26] Bouchaud J-P 2013 Crises and collective socio-economic phenomena: simple models and challenges *J. Stat. Phys.* **151** 567–606
[27] Campajola C, Lillo F and Tantari D 2020 Unveiling the relation between herding and liquidity with trader lead-lag networks *Quant. Finance* **20** 1765–78
[28] Campajola C, Di Gangi D, Lillo F and Tantari D 2020 Modelling time-varying interactions in complex systems: the score driven kinetic ising model (arXiv:2007.15545)
[29] Jaynes E T 1957 Information theory and statistical mechanics *Phys. Rev.* **106** 620
[30] Barnett L, Lizier J T, Harré M, Seth A K and Terry B 2013 Information flow in a kinetic Ising model peaks in the disordered phase *Phys. Rev. Lett.* **111** 177203
[31] Steve H *et al* 2010 Discrete temporal models of social networks *Electron. J. Stat.* **4** 585–605
[32] Park J and Newman M E J 2004 Statistical mechanics of networks *Phys. Rev.* E **70** 066117
[33] Dehnert M, Helm W E and Hütt M-T 2003 A discrete autoregressive process as a model for short-range correlations in DNA sequences *Physica* A **327** 535–53
[34] Kim J, Kim B and Sohraby K 2008 Mean queue size in a queue with discrete autoregressive arrivals of order p *Ann. Oper. Res.* **162** 69–83
[35] Williams O E, Lillo F and Latora V 2019 Effects of memory on spreading processes in non-Markovian temporal networks *New J. Phys.* **21** 043028

J. Stat. Mech. (2021) 033412

[36] Taranto D E, Bormetti G and Lillo F 2014 The adaptive nature of liquidity taking in limit order books *J. Stat. Mech.* P06002

[37] Mazzarisi P, Barucca P, Lillo F and Tantari D 2020 A dynamic network model with persistent links and node-specific latent variables, with an application to the interbank market *Eur. J. Oper. Res.* **281** 50–65

[38] Mézard M and Sakellariou J 2011 Exact mean-field inference in asymmetric kinetic Ising systems *J. Stat. Mech.* L07001

[39] Tsay R S 2014 *Financial Time Series* pp 1–23 (Wiley StatsRef: Statistics Reference Online)

[40] Decelle A and Zhang P 2015 Inference of the sparse kinetic Ising model using the decimation method *Phys. Rev.* E **91** 052136

[41] Roudi Y, Dunn B and Hertz J 2015 Multi-neuronal activity and functional connectivity in cell assemblies *Curr. Opin. Neurobiol.* **32** 38–44

J. Stat. Mech. (2021) 033412