Maximum entropy models for financial systems

Almog, A.

Appendix A

Appendix

A.1 Time series models

Models for single time series

We consider the case $N = 1$, i.e. when $\mathbf{X}$ is a $1 \times T$ matrix or equivalently a T-dimensional row vector. Let us denote the entries of $\mathbf{X}$ as $x(t)$.

Uniform random walk model

The trivial model is obtained when no constraints are enforced. In this case, there is no free parameter and the Hamiltonian has the form

$$H(\mathbf{X}) = 0 \tag{A.1}$$

As a result, the partition function is

$$Z = \sum_{\mathbf{X}} 1 = 2^T \tag{A.2}$$

which is nothing but the number of possible binary time series of length T. The probability of occurrence of a time series $\mathbf{X}$ is then

$$P(\mathbf{X}) = \frac{1}{Z} = 2^{-T} \tag{A.3}$$

and is completely uniform over the ensemble of all binary time series of length T. All the T elements of $\mathbf{X}$ are mutually independent and identically distributed with probability

$$P_t(x) \equiv \text{Prob}\big(x(t) = x\big) = \begin{cases} 1/2 & x = -1 \\ 1/2 & x = +1 \end{cases} \tag{A.4}$$

This results in a completely uniform random walk with zero expected value for each increment:

$$\langle x(t) \rangle = 0 \tag{A.5}$$

While the (ensemble) variance of each increment equals

$$\mathrm{Var}[x(t)] \equiv \langle x^2(t) \rangle - \langle x(t) \rangle^2 = 1. \tag{A.6}$$

Biased random walk model

We now consider the total increment as the simplest non-trivial (one-dimensional) constraint:

$$C(\mathbf{X}) = T \cdot M_1(\mathbf{X}) = T \cdot \overline{x(t)} \tag{A.7}$$

If we denote the corresponding (scalar) Lagrange multiplier by θ, the Hamiltonian has the form

$$H(\mathbf{X}, \theta) = \theta \cdot T \cdot \overline{x(t)} = \theta \sum_{t=1}^{T} x(t). \tag{A.8}$$

The partition function is

$$\begin{aligned} Z(\theta) &= \sum_{\mathbf{X}} e^{-\theta \sum_{t=1}^{T} x(t)} = \sum_{\mathbf{X}} \prod_{t=1}^{T} e^{-\theta x(t)} \\ &= \prod_{t=1}^{T} \sum_{x=\pm 1} e^{-\theta x} = \prod_{t=1}^{T} \left[e^{-\theta} + e^{+\theta} \right] \\ &= \left[e^{-\theta} + e^{+\theta} \right]^T \end{aligned} \tag{A.9}$$

where, when interchanging the order of the sum and product, we have replaced the sum over all time series $\mathbf{X}$ with the sum over the two possible values $x = \pm 1$ of each individual entry.

The probability of the occurrence of a time series $\mathbf{X}$ is

$$\begin{aligned} P(\mathbf{X}|\theta) &= \frac{e^{-\theta \sum_{t=1}^{T} x(t)}}{\left[e^{-\theta} + e^{+\theta} \right]^T} = \prod_{t=1}^{T} \frac{e^{-\theta x(t)}}{e^{-\theta} + e^{+\theta}} \\ &= \prod_{t=1}^{T} P_t\big(x(t)|\theta\big) \end{aligned} \tag{A.10}$$

where we have introduced the probability $P_t(x|\theta)$ of a given increment $x = \pm 1$ at time t, which we identify as

$$P_t(x|\theta) = \frac{e^{-\theta x}}{e^{-\theta} + e^{+\theta}}. \tag{A.11}$$

The above expression shows that the stochastic process corresponding to this model is a biased random walk, as the two outcomes $x = \pm 1$ have a different probability, unless $\theta = 0$ (which leads us back to the uniform random walk model considered above).

The expected value of the t-th increment $x(t)$ (representing the bias of the random walk) is

$$\langle x(t)\rangle_\theta = \sum_{x=\pm 1} x P_t(x|\theta) = \frac{e^{-\theta} - e^{+\theta}}{e^{-\theta} + e^{+\theta}} = -\tanh\theta \tag{A.12}$$

and the variance is

$$\mathrm{Var}[x(t)] = \langle x^2(t)\rangle_\theta - \langle x(t)\rangle_\theta{}^2 = 1 - \tanh^2\theta. \tag{A.13}$$

The maximum likelihood condition (1.16), fixing the value θ^* of the parameter θ given a real time series $\mathbf{X}^*$, reads

$$T\overline{\langle x(t)\rangle} = \sum_{t=1}^{T}\langle x(t)\rangle = -T\tanh\theta = T\cdot\overline{x^*(t)} \tag{A.14}$$

Where $\overline{x^*(t)}$ is the measured average increment in the observed time series $\mathbf{X}^*$. This yields

$$-\tanh\theta^* = \overline{x^*(t)} \tag{A.15}$$

which gives a parameter value

$$\theta^* = -\mathrm{artanh}\left[\overline{x^*(t)}\right] = -\frac{1}{2}\ln\left[\frac{1+\overline{x^*(t)}}{1-\overline{x^*(t)}}\right] \tag{A.16}$$

One-dimensional Ising model

We now consider a model where, besides the constraint on the total increment specified in eq.(1.33), we enforce an additional constraint on the time-delayed (lagged) quantity $T\cdot B_1(\mathbf{X})$, where $B_1(\mathbf{X})$ is defined in eq.(1.27) with $\tau = 1$. This amounts to enforce the average one-step temporal autocorrelation of the time series. The resulting 2-dimensional constraint can be written as the column vector

$$\vec{C}(\mathbf{X}) = \begin{pmatrix} C_1(\mathbf{X}) \\ C_2(\mathbf{X}) \end{pmatrix} = T\cdot\begin{pmatrix} M_1(\mathbf{X}) \\ B_1(\mathbf{X}) \end{pmatrix}. \tag{A.17}$$

If we write the corresponding Lagrange multiplier as

$$\vec{\theta} = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} = -\begin{pmatrix} I \\ K \end{pmatrix}, \tag{A.18}$$

then the Hamiltonian reads

$$\begin{aligned} H(\mathbf{X}, I, K) &= \vec{\theta} \cdot \vec{C}(\mathbf{X}) = T\theta_1 M_1(\mathbf{X}) + T\theta_2 B_1(\mathbf{X}) \\ &= -I \sum_{t=1}^{T} x(t) - K \sum_{t=1}^{T} x(t)x(t+1), \end{aligned} \tag{A.19}$$

where we consider a periodicity condition as in eq.(1.28) with $\tau = 1$, i.e. $x(T+1) \equiv x(1)$. Note that, when $\mathbf{X}$ is a real binary time series of length T, this condition can be always enforced by adding one last (fictious) timestep $T+1$ and a corresponding increment $x(T+1)$ chosen equal to $x(1)$. For long time series, this has a negligible effect.

The above Hamiltonian coincides with that for the one-dimensional Ising model with periodic boundary conditions [55] (chapter 1). Each time step t is seen as a site in an ordered chain of length T, and each value $x(t) = \pm 1$ is seen as the value of a spin sitting at that site. The model is analytically solvable, which allows us to apply it to real time series in our formalism. For the readers familiar with time series analysis but not necessarily with the Ising model, we briefly recall the standard solution of the model, adapting it from ref. [55] (chapter 1).

Applying the periodicity condition of eq.(1.28) ensures that all sites (time steps) are statistically equivalent, i.e.:

$$\langle x(1) \rangle = \langle x(2) \rangle = \cdots = \langle x(T) \rangle \tag{A.20}$$

so that the system is translationally (here, temporally) invariant. The partition function is

$$Z(I, K) = \sum_{\mathbf{X}} \exp \left[I \sum_{t=1}^{T} x(t) + K \sum_{t=1}^{T} x(t)x(t+1) \right]$$

and can be rewritten as a product of terms involving only two successive time steps:

$$Z(I, K) = \sum_{\mathbf{X}} \prod_{t=1}^{T} V\big(x(t), x(t+1)\big), \tag{A.21}$$

where we have introduced the function $V(x, y)$ defined as

$$V(x, y) \equiv \exp \left(I \frac{x+y}{2} + Kxy \right). \tag{A.22}$$

We since both x and y can take only the values ± 1, we can regard $V(x, y)$ as the element of a 2×2 matrix $\mathbf{V}$ called the *transfer matrix* [55] (chapter 1):

$$\mathbf{V} \equiv \begin{pmatrix} V(+1,+1) & V(+1,-1) \\ V(-1,+1) & V(-1,-1) \end{pmatrix} = \begin{pmatrix} e^{K+I} & e^{-K} \\ e^{-K} & e^{K-I} \end{pmatrix}. \tag{A.23}$$

This allows us to rewrite eq.(A.21) as

$$Z(I,K) = \mathrm{Tr}\big(\mathbf{V}^T\big). \tag{A.24}$$

Let $\vec{v}_1, \vec{v}_2$ denote the two eigenvectors of $\mathbf{V}$, and λ_1, λ_2 the corresponding eigenvalues, so that

$$\mathbf{V}\vec{v}_j = \lambda_j \vec{v}_j, \quad j = 1, 2. \tag{A.25}$$

The 2×2 matrix $\mathbf{Q} \equiv (\vec{v}_1, \vec{v}_2)$ (having column vectors $\vec{v}_1$ and $\vec{v}_2$) diagonalizes $\mathbf{V}$, i.e.

$$\mathbf{V} = \mathbf{Q} \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \mathbf{Q}^{-1}, \tag{A.26}$$

where a direct calculation of the eigenvalues and eigenvectors yields

$$\lambda_1 = e^K \cosh I + \sqrt{e^{2K} \sinh^2 I + e^{-2K}} \tag{A.27}$$

$$\lambda_2 = e^K \cosh I - \sqrt{e^{2K} \sinh^2 I + e^{-2K}} \tag{A.28}$$

and

$$\mathbf{Q} = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix}, \tag{A.29}$$

with ϕ defined by

$$\cot 2\phi \equiv e^{2K} \sinh I. \tag{A.30}$$

It then follows that eq.(A.24) simply reduces to

$$Z(I,K) = \mathrm{Tr} \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}^T = \lambda_1^T + \lambda_2^T, \tag{A.31}$$

and the probability of occurrence of a time series $\mathbf{X}$ is

$$P(\mathbf{X}|I,K) = \frac{\prod_{t=1}^{T} V\big(x(t), x(t+1)\big)}{\lambda_1^T + \lambda_2^T}. \tag{A.32}$$

The above results allow us to analytically obtain expected values. That of $x(t)$ is

$$\langle x(t) \rangle = \sum_{\mathbf{X}} x(t) P(\mathbf{X}|I,K) = \frac{\mathrm{Tr}\big(\mathbf{S}\mathbf{V}^T\big)}{\lambda_1^T + \lambda_2^T}, \tag{A.33}$$

where we have introduced the diagonal matrix

$$\mathbf{S} \equiv \begin{pmatrix} S(+1,+1) & S(+1,-1) \\ S(-1,+1) & S(-1,-1) \end{pmatrix} = \begin{pmatrix} +1 & 0 \\ 0 & -1 \end{pmatrix} \tag{A.34}$$

having elements

$$S(x,y) \equiv x\delta(x,y). \tag{A.35}$$

Similarly, for $0 < s-t < T$ the expected value of $x(t)x(s)$ is

$$\begin{aligned}\langle x(t)x(s)\rangle &= \sum_{\mathbf{X}} x(t)x(s)P(\mathbf{X}|I,K) \\ &= \frac{\mathrm{Tr}\left(\mathbf{S}\mathbf{V}^{s-t}\mathbf{S}\mathbf{V}^{T+t-s}\right)}{\lambda_1^T+\lambda_2^T}.\end{aligned} \tag{A.36}$$

In the limit $T \to \infty$ (corresponding to long time series in our case) with $s-t$ fixed, these expressions become

$$\langle x(t)\rangle = \cos 2\phi \tag{A.37}$$

$$\langle x(t)x(s)\rangle = \cos^2 2\phi + \sin^2 2\phi \left(\frac{\lambda_1}{\lambda_2}\right)^{s-t} \tag{A.38}$$

Now, we note that eqs.(A.33) and (A.36) manifestly show the translational (temporal) invariance of the model, as $\langle x(t)\rangle$ is independent of t and $\langle x(t)x(s)\rangle$ depends on t and s only through their difference $s-t$. This implies that, writing $\tau \equiv s-t$ and performing a temporal average,

$$\left\langle M_1 \right\rangle = \cos 2\phi \tag{A.39}$$

$$\left\langle B_\tau \right\rangle = \cos^2 2\phi + \sin^2 2\phi \left(\frac{\lambda_1}{\lambda_2}\right)^{\tau}. \tag{A.40}$$

Using eq.(A.30) we can rewrite these expressions in terms of the model parameters, I and K, as

$$\left\langle M_1 \right\rangle = \frac{e^{2K}\sinh I}{\sqrt{1+e^{4K}\sinh^2 I}} \tag{A.41}$$

$$\left\langle B_\tau \right\rangle = \frac{e^{4K}\sinh^2 I + (\lambda_1/\lambda_2)^\tau}{1+e^{4K}\sinh^2 I}. \tag{A.42}$$

The expected value of the autocorrelation defined in eq. (1.47) can be approximated as the ratio of two expected values as follows:

$$\left\langle A_\tau \right\rangle \equiv \left\langle \frac{B_\tau - M_1^2}{1-M_1^2} \right\rangle \approx \frac{\langle B_\tau\rangle - \langle M_1^2\rangle}{1-\langle M_1^2\rangle} = \left(\frac{\lambda_1}{\lambda_2}\right)^{\tau}. \tag{A.43}$$

Models for single cross-sections of multiple time series

For a single cross-section of a set of N multiple time series, $\mathbf{X}$ is a $N \times 1$ matrix or equivalently a N-dimensional column vector. We denote the entries of $\mathbf{X}$ as x_i.

Uniform random walk model

The uniform random walk is a simple modification of the same model that we considered for single time series, where $x(t)$ is replaced by x_i and T is replaced by N. This model is obtained when no constraints are enforced. The Hamiltonian is

$$H(\mathbf{X}) = 0 \tag{A.44}$$

and the partition function is simply the number of possible configurations for a single cross-section of N stocks:

$$Z = \sum_{\mathbf{X}} 1 = 2^N. \tag{A.45}$$

The probability of occurrence of a cross section $\mathbf{X}$ is

$$P(\mathbf{X}) = \frac{1}{Z} = 2^{-N} \tag{A.46}$$

and is completely uniform over the ensemble of all cross sections of N stocks. All the N elements of $\mathbf{X}$ are mutually independent and identically distributed with probability

$$P_i(x) \equiv \text{Prob}\big(x_i = x\big) = \begin{cases} 1/2 & x = -1 \\ 1/2 & x = +1 \end{cases} \tag{A.47}$$

This results in a completely uniform random walk with zero expected value

$$\langle x_i \rangle = 0 \tag{A.48}$$

and maximum variance

$$\text{Var}[x_i] \equiv \langle x_i^2 \rangle - \langle x_i \rangle^2 = 1. \tag{A.49}$$

Biased random walk model

Also this model is analogous to the corresponding model for single time series. We select the total daily increment of the cross section $\mathbf{X}$ as the constraint:

$$C(\mathbf{X}) = N \cdot M_1(\mathbf{X}) = N \cdot \{x_i\} \tag{A.50}$$

Let the corresponding Lagrange multiplier be denoted by θ. The Hamiltonian is

$$H(\mathbf{X}, \theta) = \theta \cdot N \cdot \{x_i\} = \theta \sum_{i=1}^{N} x_i \tag{A.51}$$

and the partition function is

$$\begin{aligned} Z(\theta) &= \sum_{\mathbf{X}} e^{-\theta \sum_{i=1}^{N} x_i} = \sum_{\mathbf{X}} \prod_{i=1}^{N} e^{-\theta x_i} \\ &= \prod_{i=1}^{N} \sum_{x=\pm 1} e^{-\theta x} = \prod_{i=1}^{N} \left[e^{-\theta} + e^{+\theta}\right] \\ &= \left[e^{-\theta} + e^{+\theta}\right]^N . \end{aligned} \tag{A.52}$$

The probability of the occurrence of a cross section $\mathbf{X}$ is

$$\begin{aligned} P(\mathbf{X}|\theta) &= \frac{e^{-\theta \sum_{i=1}^{N} x_i}}{\left[e^{-\theta} + e^{+\theta}\right]^N} = \prod_{i=1}^{N} \frac{e^{-\theta x_i}}{e^{-\theta} + e^{+\theta}} \\ &= \prod_{i=1}^{N} P_i\big(x_i|\theta\big) \end{aligned} \tag{A.53}$$

where we have introduced the probability $P_i(x|\theta)$ of a given increment $x = \pm 1$ for stock i, which we identify as

$$P_i(x|\theta) = \frac{e^{-\theta x}}{e^{-\theta} + e^{+\theta}}. \tag{A.54}$$

Just like the corresponding model for single time series, this model is a biased random walk, because the two outcomes $x = \pm 1$ have a different probability unless $\theta = 0$.

The expected value of the i-th increment x_i is

$$\langle x_i \rangle_\theta = \sum_{x=\pm 1} x P_i(x|\theta) = \frac{e^{-\theta} - e^{+\theta}}{e^{-\theta} + e^{+\theta}} = -\tanh\theta \tag{A.55}$$

and the variance is

$$\mathrm{Var}[x_i] = \langle x_i^2 \rangle_\theta - \langle x_i \rangle_\theta{}^2 = 1 - \tanh^2\theta. \tag{A.56}$$

The maximum likelihood condition (1.16), fixing the value θ^* of the parameter θ given a real cross section $\mathbf{X}^*$, reads

$$N\big\langle \{x_i\} \big\rangle = \sum_{i=1}^{N} \langle x_i \rangle = -N \tanh\theta = N \cdot \{x_i^*\} \tag{A.57}$$

where $\{x_i^*\}$ is the measured average increment of the observed cross section $\mathbf{X}^*$. This yields

$$-\tanh\theta^* = \{x_i^*\} \tag{A.58}$$

which gives a parameter value

$$\theta^* = -\mathrm{artanh}\left[\{x_i^*\}\right] = -\frac{1}{2} \ln\left[\frac{1 + \{x_i^*\}}{1 - \{x_i^*\}}\right]. \tag{A.59}$$

Mean-field Ising model

In this model, we enforce two constraints: the total increment and the total coupling between stocks. The resulting 2-dimensional constraint can be written as

$$\vec{C}(\mathbf{X}) = \begin{pmatrix} C_1(\mathbf{X}) \\ C_2(\mathbf{X}) \end{pmatrix} = \begin{pmatrix} N \cdot M_1(\mathbf{X}) \\ D(\mathbf{X}) \end{pmatrix}. \tag{A.60}$$

We can write the corresponding Lagrange multiplier as

$$\vec{\theta} = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} = - \begin{pmatrix} h \\ J \end{pmatrix} \tag{A.61}$$

and the Hamiltonian as

$$H(\mathbf{X}, h, J) = -h \sum_{i=1}^{N} x_i - J \sum_{i<j} x_i x_j. \tag{A.62}$$

Note that here we are not enforcing nearest-neighbor interactions as in the one-lagged model for single time series, but market-wide interactions among all stocks for the same time step (cross section). This is the result of the fact that, when considering cross sections, there is no natural notion of 'lattice sites' induced by e.g. a temporal ordering as in the one-lagged model. In other words, pairs of stocks in a cross section are neither 'close' nor 'distant'. We therefore assume a common interaction strength J among all stocks.

The above model, known as the mean-field Ising model, is analytically solvable. Here we adapt the derivation illustrated in ref. [55] (chapter 1). We first note that, since ${x_i}^2 = 1$ for all i, $H(\mathbf{X}, h, J)$ can be expressed as a function of $M_1(\mathbf{X})$ alone:

$$H(\mathbf{X}, h, J) = -hNM_1(\mathbf{X}) - \frac{J}{2}\big[N^2 M_1^2(\mathbf{X}) - N\big]. \tag{A.63}$$

This implies that the sum over configurations in the partition function can be replaced by a sum over the allowed values of $M_1(\mathbf{X})$, weighted by the number of configurations for each value. If we denote by r the number of increments that are negative ($x = -1$), and by $(N - r)$ the number of increments that are positive ($x = +1$), then we can write the Hamiltonian as a function of r alone through the expression

$$NM_1(\mathbf{X}) = N - 2r. \tag{A.64}$$

The partition function can therefore be calculated as

$$Z(h, J) \equiv \sum_{\mathbf{X}} e^{-H(\mathbf{X}, h, J)} = \sum_{r=1}^{N} C_r \tag{A.65}$$

where

$$C_r \equiv \frac{N!}{r!(N-r)!} e^{h(N-2r)+\frac{J}{2}[(N-2r)^2-N]} \tag{A.66}$$

incorporates the binomial coefficient enumerating the configurations with given r. The expected increment is therefore

$$\langle M_1 \rangle = \left\langle 1 - \frac{2r}{N} \right\rangle = \frac{\sum_{r=1}^{N} \left(1 - \frac{2r}{N}\right) C_r}{Z(h,J)} \quad \forall i. \tag{A.67}$$

When N is large, a traditional derivation [55] (chapter 1) shows that the sum at the numerator of eq.(A.67) is dominated by the single addendum corresponding to the maximum of C_r. The same applies to the partition function at the denominator. If r_0 denotes the value of r such that C_r is maximum, we then get

$$\langle M_1 \rangle \approx 1 - \frac{2r_0}{N}. \tag{A.68}$$

A further expansion [55] (chapter 1) finally shows that, given h and J, the expected value $\langle M_1 \rangle$ is the solution of the nonlinear equation

$$\langle M_1 \rangle = \tanh\left[(N-1)J\langle M_1 \rangle + h\right]. \tag{A.69}$$

From the above equation, one can infer the existence of a phase transition in the model, separating a regime where the expected 'magnetization' (here the average increment $\langle M_1 \rangle$) is zero from one where it is non-zero [55] (chapter 1). This transition is discussed in sec.1.5.3.

Before proceeding further, we note a peculiarity of the model, which has implications for the applicability of our maximum likelihood approach. An argument similar to that leading to eq.(A.68) implies that the second moment of $M_1(\mathbf{X})$ can be expressed as

$$\langle M_1^2 \rangle = \left\langle \left(1 - \frac{2r}{N}\right)^2 \right\rangle \approx \left(1 - \frac{2r_0}{N}\right)^2 \approx \langle M_1 \rangle^2. \tag{A.70}$$

This implies that

$$\mathrm{Var}[M_1] \equiv \langle M_1^2 \rangle - \langle M_1 \rangle^2 = 0, \tag{A.71}$$

or in other words that $M_1(\mathbf{X})$ is no longer a random variable. As a consequence, something unusual happens when we apply the maximum likelihood principle. From eq.(A.63), and recalling the general result embodied by eq.(1.20) in sec.1.3.2, it is clear that the parameter values h^* and J^* maximizing the likelihood can be found as the solution to the two coupled equations

$$\langle M_1 \rangle = M_1(\mathbf{X}^*) \tag{A.72}$$

$$\langle M_1^2 \rangle = M_1^2(\mathbf{X}^*) \tag{A.73}$$

However, eq. (A.70) implies that eq.(A.73) can be rewritten as

$$\langle M_1 \rangle^2 = M_1^2(\mathbf{X}^*) \tag{A.74}$$

which coincides with eq.(A.72). So eqs. (A.72) and (A.73) are equivalent, and they cannot be used to uniquely determine the two unknown parameters h^* and J^*. This is the result of the fact that, when fitted to the data, the model is actually over-constrained: there are two parameters to fit the only constraint (M_1) on which the Hamiltonian depends. This aspect of the model is not manifest when M_1 is regarded as a function of h and J, as usually done when simulating spin systems.

The above consideration implies that we should drop one of the two parameters and consider the two cases $J = 0$ and $h = 0$ separately. The former case coincides with the biased random walk model that we already discussed, and we will not discuss it any further. The latter case will instead represent our genuine specification of the 'mean-field' model. Setting $h = 0$ implies

$$H(\mathbf{X}, 0, J) = -\frac{J}{2}\big[N^2 M_1^2(\mathbf{X}) - N\big] \tag{A.75}$$

and

$$\langle M_1 \rangle = \tanh\big[(N-1)J\langle M_1 \rangle\big]. \tag{A.76}$$

Applying the maximum likelihood principle to eq.(A.75) tells us to select J^* as the solution of eq.(A.73). However, we have seen that this condition leads to eq.(A.74), which is actually equivalent to eq.(A.72). Therefore, the value of J^* can be found by replacing $\langle M_1 \rangle$ with the observed value $M_1(\mathbf{X}^*) = \{x_i^*\}$ in eq. (A.76), which leads to

$$\{x_i^*\} = \tanh\big[(N-1)J^*\{x_i^*\}\big]. \tag{A.77}$$

Note that in the traditional situation one is interested in finding the (expected) magnetization given a value of J, which implies that the transcendental eq. (A.76) should be solved numerically. Here, we are instead facing the inverse situation where we look for the value of J^* given the (observed) value of the magnetization. In this quite unusual case, it turns out that eq. (A.77) can be inverted to give the following analytical solution:

$$J^* = \frac{\operatorname{artanh}\{x_i^*\}}{\{x_i^*\}(N-1)} = \frac{1}{2\{x_i^*\}(N-1)} \ln\left[\frac{1+\{x_i^*\}}{1-\{x_i^*\}}\right]. \tag{A.78}$$

Once this value is calculated, it can be inserted into the probability

$$P(\mathbf{X}^*|0, J) = \frac{e^{-H(\mathbf{X}^*,0,J)}}{Z(0,J)} = \frac{e^{JN(N\{x_i^*\}^2-1)/2}}{\sum_{r=1}^{N} \frac{N!}{r!(N-r)!} e^{J[(N-2r)^2-N]/2}} \tag{A.79}$$

(where we have set $h = 0$) to obtain the maximized likelihood of generating the observed cross section $\mathbf{X}^*$ under the mean-field model.

In this appendix we give a summarized description of the binary and weighted network quantities which are studied in this paper. Specifically, we first show how the properties are measured over a real network, and then how the expected values under the ECM and the TS model are constructed.

A.2 Network models

Observed Properties

Let us note a weighted undirected network as a square matrix $\mathbf{W}$, where the specific entry w_{ij} represents the edge weight between country i and country j. The binary representation of the network is noted by a binary matrix $\mathbf{A}$, where the entries are $a_{ij} = \Theta[w_{ij}]$, $\forall\, i, j$.

We compute the Average Nearest Neighbor Degree as:

$$k_i^{nn}(\mathbf{W}) = \sum_{j \neq i} \frac{a_{ij} k_j}{k_i} = \frac{\sum_{j \neq i} \sum_{k \neq j} a_{ij} a_{jk}}{\sum_{j \neq i} a_{ij}}. \tag{A.80}$$

Its calculated as the arithmetic mean of the degrees of the neighbors of a specific node, which is a measure of correlation between the degrees of adjacent nodes.

The Binary Clustering Coefficient has the following expression:

$$c_i(\mathbf{W}) = \frac{\sum_{j \neq i} \sum_{k \neq i,j} a_{ij} a_{jk} a_{ki}}{\sum_{j \neq i} \sum_{k \neq i,j} a_{ij} a_{ki}}. \tag{A.81}$$

It is a measure of the tendency to which nodes in a graph form cluster together. More specifically, it counts how many closed triangles are attached to each node with respect to all the possible triangles.

The corresponding weighted properties are the Average Nearest Neighbor Strength and the weighted Clustering Coefficient. The Average Nearest Neighbor Strength, defined as:

$$s_i^{nn}(\mathbf{W}) = \sum_{j \neq i} \frac{a_{ij} s_j}{k_i} = \frac{\sum_{j \neq i} \sum_{k \neq j} a_{ij} w_{jk}}{\sum_{j \neq i} a_{ij}} \tag{A.82}$$

where $s_i = \sum_j w_{ij}$ is the strength (total flow) of a country. The s_i^{nn} measure the average strength of the neighbors for a specific node i. Like its binary counterpart, it gives the magnitude of activity of a specific node neighbors (weighted activity).

The weighted Clustering Coefficient [54] (chapter 2) is defined as:

$$c_i^W(\mathbf{W}) = \frac{\sum_{j\neq i}\sum_{k\neq i,j}(w_{ij}w_{jk}w_{ki})^{\frac{1}{3}}}{\sum_{j\neq i}\sum_{k\neq i,j}a_{ij}a_{ki}}. \tag{A.83}$$

The $c_i^W(\mathbf{W})$ is a measure of the weight density in the neighborhood of a node. It classify the tendency of a specific node to cluster in a triangle taking into account also the edge-values.

Now, the measured properties of the real network need to be compared with the reproduced properties of the different models. These reproduced properties are the expected values of the maximum entropy ensemble that each model id generating, and can be calculated analytically. The expected values can be obtained by simply replacing a_ij with the probability p_{ij} for the different models (p_{ij} is different to each model). This next step is what we will discuss in the next sections.

Expected values in the BCM and ECM

Since the BCM model is only dealing with the binary representation, we will have expected values just for the two binary higher-order properties. While the ECM gives expectations for the weighted counterparts of the binary properties.

For the binary higher-order properties, we replace a_ij with p_{ij} which is the probability of creating a link, and also the expected value of the edge $p_{ij} = \langle a_{ij}\rangle$. This simple procedure yields the analytic formula of the expected value for the properties. We compute the expected Average Nearest Neighbor Degree as:

$$\langle k_i^{nn}\rangle = \frac{\sum_{j\neq i}\sum_{k\neq j}p_{ij}p_{jk}}{\sum_{j\neq i}p_{ij}} \tag{A.84}$$

and the expected Binary Clustering Coefficient as:

$$\langle c_i\rangle = \frac{\sum_{j\neq i}\sum_{k\neq i,j}p_{ij}p_{jk}p_{ki}}{\sum_{j\neq i}\sum_{k\neq i,j}p_{ij}p_{ki}} \tag{A.85}$$

where for the BCM model we input $p_{ij} = \frac{z_i z_j}{1+z_i z_j}$, and for the ECM the more complex term $p_{ij} = \frac{x_i x_j y_i y_j}{1-y_i y_j + x_i x_j y_i y_j}$

In the weighted case (weighted higher-order properties), we are left only with the ECM. The expected Average Nearest Neighbor Strength is calculated as:

$$\langle s_i^{nn}\rangle = \frac{\sum_{j\neq i}\sum_{k\neq j}p_{ij}\langle w_{jk}\rangle}{\sum_{j\neq i}p_{ij}} \tag{A.86}$$

where $\langle w_{jk} \rangle = \frac{x_i x_j y_i y_j}{(1-y_i y_j + x_i x_j y_i y_j)(1-y_i y_j)}$ and we input the p_{ij} of the ECM model as before.

In the expected value of the c^W we should be more careful, since it is necessary to calculate the expected product of (powers of) distinct matrix entries

$$\langle c_i^W \rangle = \frac{\sum_{j \neq i} \sum_{k \neq i,j} \langle (w_{ij} w_{jk} w_{ki})^{\frac{1}{3}} \rangle}{\sum_{j \neq i} \sum_{k \neq i,j} p_{ij} p_{ki}}. \tag{A.87}$$

We know that

$$\langle \sum_{i \neq j \neq k} w_{ij}^{\alpha} \cdot w_{jk}^{\beta} \cdot ... \rangle = \sum_{i \neq j \neq k} \langle w_{ij}^{\alpha} \rangle \cdot \langle w_{jk}^{\beta} \rangle \cdot \langle ... \rangle \tag{A.88}$$

with the generic term for the ECM case

$$\langle w_{ij}^{\gamma} \rangle = \sum_{0}^{\infty} w^{\gamma} q_{ij}(w|\vec{x}, \vec{y}) = \frac{x_i x_j (1 - y_i y_j) Li_{\gamma}(y_i y_j)}{1 - y_i y_j + x_i x_j y_i y_j} \tag{A.89}$$

where $Li_n(R) = \sum_{l=1}^{\infty} \frac{R^l}{l^n}$ is the nth polylogarithm of R. For a more comprehensive description please refer to [32] (chapter 2).

Expected values in the TS model

Here again we use the known expressions for the properties and replacing the terms p_{ij}^{ts} and w_{ij}^{ts} with the expected values $\langle a_{ij} \rangle$ and $\langle w_{ij} \rangle$ correspondingly. However, here the expected values are a function of the GDP of the countries, or more specifically the re-scaled GDP g_i. The expressions for the higher-order binary properties are as before :

$$\langle k_i^{nn} \rangle = \frac{\sum_{j \neq i} \sum_{k \neq j} p_{ij}^{ts} p_{jk}^{ts}}{\sum_{j \neq i} p_{ij}^{ts}} \tag{A.90}$$

and

$$\langle c_i \rangle = \frac{\sum_{j \neq i} \sum_{k \neq i,j} p_{ij}^{ts} p_{jk}^{ts} p_{ki}^{ts}}{\sum_{j \neq i} \sum_{k \neq i,j} p_{ij}^{ts} p_{ki}^{ts}} \tag{A.91}$$

where $p_{ij}^{ts} = \frac{a g_i g_j}{1 + a g_i g_j}$.

In the weighted case,the expected Average Nearest Neighbor Strength is calculated as:

$$\langle s_i^{nn} \rangle = \frac{\sum_{j \neq i} \sum_{k \neq j} p_{ij}^{ts} \langle w_{jk}^{ts} \rangle}{\sum_{j \neq i} p_{ij}^{ts}} \tag{A.92}$$

where $\langle w_{ij}\rangle^{ts} = \frac{ag_ig_j}{1+ag_ig_j}\cdot\frac{(1+bg_i^c)(1+bg_j^c)}{(1+bg_i^c+bg_j^c)}$ and we input the p_{ij}^{ts} of the Two-Step model as before.

For convenience reasons we will write the expression for the weighted Clustering Coefficient c^W first as a function of the fitness parameters z_i and y_i, and later replaced them with the corresponding GDP terms. The expected value of the c^W in the Two-Step case is :

$$\langle c_i^W\rangle = \frac{\sum_{j\neq i}\sum_{k\neq i,j}\langle (w_{ij}w_{jk}w_{ki})^{\frac{1}{3}}\rangle^{ts}}{\sum_{j\neq i}\sum_{k\neq i,j}p_{ij}^{ts}p_{ki}^{ts}}. \tag{A.93}$$

As before we observe that

$$\langle \sum_{i\neq j\neq k} w_{ij}^{\alpha}\cdot w_{jk}^{\beta}\cdot ...\rangle = \sum_{i\neq j\neq k}\langle w_{ij}^{\alpha}\rangle\cdot\langle w_{jk}^{\beta}\rangle\cdot\langle ...\rangle \tag{A.94}$$

with the generic term for the Two-Step model

$$\langle w_{ij}^{\gamma}\rangle = \sum_0^{\infty} w^{\gamma}q_{ij}(w|\vec{z},\vec{y}) = \frac{z_iz_j(1-y_iy_j)Li_{\gamma}(y_iy_j)}{(1+z_iz_j)y_iy_j} \tag{A.95}$$

where $Li_n(R) = \sum_{l=1}^{\infty}\frac{R^l}{l^n}$ is the nth polylogarithm of R.

Once we input the expressions of z_i and y_i

$$\begin{aligned} z_i &= \sqrt{a}\cdot g_i, \\ y_i &= \frac{b\cdot g_i^c}{1+b\cdot g_i^c} \end{aligned} \tag{A.96}$$

equation (A.95) yields

$$\langle w_{ij}^{\gamma}\rangle = \frac{ag_ig_j}{1+ag_ig_j}\cdot\frac{1+cg_i^c+bg_j^c}{b^2g_i^cg_j^c}Li_{\gamma}\left(\frac{b^2g_i^cg_j^c}{(1+bg_i^c)(1+bg_j^c)}\right) \tag{A.97}$$