The Risk of Random Sets with Applications to Basket Derivatives

Christian Gouriéroux[1] Yang Lu[2] Alain Monfort[3]

October 2023. Preliminary version, all comments welcome.

Abstract: This paper analyzes the risks in random sets and their implications for basket derivatives. Based on an extension of integration by parts for random set, we define stochastic dominance of order 1 and 2 for random sets. Since the ordering of sets, that is the inclusion, is a partial order, we have to distinguish left and right notions of stochastic dominance. The observed sets are in a one-to-one relationship with observed multivariate binary variables, each component of which indicating high or low risk for a given type of risk. This relationship is used to define basket derivatives and to develop statistical inference. We consider the special cases of exchangeability, of Law of Determinantal Point Process (LDPP), of local pairwise interactions and of block models for illustration.

Keywords: Random Set, Multivariate Risk, Stochastic Dominance, Determinantal Point Process, Basket Derivative, Set-European Call, Exchangeability, Cyber Risk.

Acknowledgment: Gouriéroux gratefully acknowledges financial support of the ACPR Chair "Regulation and Systemic Risks", and the ERC DYSMOIA. Lu acknowledges support from NSERC through grant RGPIN-2021-04144.

[1]University of Toronto, Toulouse School of Economics and CREST. Email: gouriero@ensae.fr
[2]Concordia University, Montreal, Canada. Email: yang.lu@concordia.ca
[3]CREST. Email: alain.monfort@ensae.fr

1 Introduction

The aim of this paper is to analyze the risks in random sets and their implication for pricing baskets of individual risks. The analysis of risk is usually done by using the notion of stochastic dominance at order 1 or 2. For distributions on the real line, the stochastic dominance at order 1 (resp. order 2) involves all increasing (resp. increasing convex) functions. This has led to the introduction of special families of derivatives such as the European calls (resp. the digital tranches), which are insurance coverages against large risk at order 2 (resp. at order 1).

This basic risk theory is developed for one dimensional continuous risk variables [Hanoch and Levy (1969), Rothschild and Stiglitz (1970)[1], Vickson (1975), Fishburn and Vickson (1978)]. It has been extended to multidimensional continuous risk variables by Kihlstrom and Mirman (1974), Scarsini (1988), Marshall (1991).

We review in Section 2 different results on random sets in the finite space $\{1, ..., n\}$. We first recall the link between the observation of a set of n binary variables and a subset S of $\{1, ..., n\}$. Then we introduce notions of increasing, decreasing, decreasing convex functions of a set, by extending the standard notions of cumulative distribution and survival functions, and derive an "integration by parts" formula for such functions. We also discuss the case of a particular parametric family of set distributions: the Law of Determinantal Point Process (LDPP) family. Since the literature on random sets, LDPP and machine learning has not well diffused among researchers in economics, finance and insurance, we provide detailed references for all the known results used in the paper. All other results are new and are systematically proven.

Stochastic dominances on random sets are introduced in Section 3. We especially discuss right stochastic dominance at orders 1 and 2. In the LDPP framework, the right stochastic set dominance is equivalent to a new ordering on appropriate symmetric positive semi-definite matrices. We also define set (i.e. basket) derivatives, such as tranches and European calls, and explain how stochastic set dominance can be written in terms of the expected payoffs of basket derivatives. Section 4 discusses the case of exchangeable models where the set distributions are invariant by permutation of the individual indexes. We highlight the simplifications that arise in this case.

Illustrations are provided in Section 5. Section 6 concludes. The proofs of propositions, the derivation of Laplace transforms, different properties of the LDPP family, and of the log-linear model with pairwise interactions are provided in the Appendix and online Appendices. We also provide a basic introduction to the statistical inference of random set models in online appendix 3.

[1] With the correction in Leshno et al. (1997).

2 Random sets

The theory of random sets was introduced by Debreu (1967) and has recently also received much attention in econometrics [see e.g. Beresteanu et al. (2011), Molchanov and Molinari (2018)]. In this section, we briefly review some basic properties of random sets in finite spaces[2] and of their distributions that will be used later on for risk analysis.

2.1 Multivariate binary variables

Let us consider a sequence of doubly indexed binary variables $X_{l,i}, l = 1, ..., L, i = 1, ..., n,$, that take values 0 or 1. Here n denotes the number of individuals, and depending on the application, the interpretation of the other index might differ. In some cases, L denotes the number of binary variables for each individual, and the simplest baseline model in this case assumes that the individuals are independent, whereas the L binary variables could be dependent for the same individual. In other words, we have multivariate Bernoulli data and the focus is on within-dependence. In some other cases, each individual has only one type of binary risk, but these binary variables are time-varying and are observed during L periods. The simplest baseline model in this case assumes that the variables at different periods are independent, whereas at the same period, the different individuals are dependent. In other words, we have panel Bernoulli data and the primary focus is on between-dependence.

Because our interest is in risk analysis, in the applications below, by convention, $X_{l,i} = 1$, if the individual has a high l risk, $X_{l,i} = 0$, otherwise.

Example 1 (Multivariate binary data)**.** The individual i can be a corporate, and for different l, the variables $X_{l,i}$ can correspond to the solvency risk, the liquidity risk, or the cyber risk in a given year.

Example 2. The machine learning approaches are using large databases that contain a lot of missing data usually treated by ad-hoc imputation methods. These imputations can impact significantly the machine learning and lead to unreasonable results [see e.g. Bryzgalova et al. (2022)]. Typically the firms, or the hedge funds have to satisfy transparency requirements by disclosing a list of characteristics on their balance sheets, or on their efforts for diminishing carbon impact. The proportion of missing data is between 20 % and 50 %, depending on the firm, on the variable (and on the period). In this example, we observe $X_{ljt} = 1$, if the value of variable l is missing in the reporting by firm (hedge fund) j at date t. In this example we have $i = (j, t)$.

[2]See e.g. Matheron (1975), Molchanov (2017), Molchanov and Molinari (2018) for random sets in more general spaces.

The observation of $X = (X_{l,i}, l = 1, ..., L, i = 1, ...n)$ is equivalent to the observation of a subset $S = [S(1), ..., S(L)]$ of the product space $\left[\mathcal{P}(\{1, 2, ..., n\})\right]^L$, where for each l, the set $S(l)$ denotes the subset of individuals i with $X_{l,i} = 1$ and $\mathcal{P}(\{1, 2, ..., n\})$ is the set of all subsets of $\{1, 2, ..., n\}$. Therefore it is equivalent to analyze X (resp. a function of X), or to analyze the set S (resp. a function of S). We follow this interpretation in terms of sets in the rest of the paper.

When the binary variables are doubly indexed by individual i and time t, they can be transformed into observations of sets in three different ways, that are:

a) subsets of $\{1, ..., n\} \times \{1, ..., L\}$ indexed by time;

b) subsets of $\{1, ..., T\} \times \{1, ..., L\}$ indexed by individual;

c) subsets of $\{1, ..., L\}$ indexed by individual and time.

The choice between $a), b)$ and $c)$ will depend on the type of application.

As an illustration, let us consider $n = 6$, $L = 3$ and the following individual observations of variables $X_{1,i}$, $X_{2,i}$, and $X_{3,i}$, $i = 1, 2, ..., 6$, the associated S_i and its size $n(S_i)$, $i = 1, 2, ..., 6$, as well as $S(l)$ along with its size $n[S(l)]$, $l = 1, 2, 3,$. As in Example 1, each column is an individual (corporate), whereas each row corresponds to one type of risk (solvency, liquidity and cyber). In particular, in the event that the simultaneous occurrence of $X_{1,i} = 1$, $X_{2,i} = 1$ and $X_{3,i} = 1$ is rare, the representation through set could require less computer memory.

l \ i	1	2	3	4	5	6	$S(l)$	$n[S(l)]$
$l = 1$	0	0	0	1	1	1	$\{4, 5, 6\}$	3
$l = 2$	1	0	0	0	0	1	$\{1, 6\}$	2
$l = 3$	0	1	0	0	0	0	$\{2\}$	1
S_i	$\{2\}$	$\{3\}$	$\emptyset$	$\{1\}$	$\{1\}$	$\{1, 2\}$		
$n(S_i)$	1	1	0	1	1	2		

Table 1: Two equivalent representations of the doubly indexed binary variables: through a matrix of binary entries $X_{l,i}, l = 1, 2, 3, i = 1, .., 6$, or through a sequence of set-valued variables $S_i, i = 1, ..., 6$, or $S(l), l = 1, ..., 3$.

2.2 Functions of sets

The inclusion (in the wide sense including equality) defines a partial order[3] on all these sets[4] $s = [s(1), ..., s(L)]$ with extremal elements: the empty set $\emptyset$ as the minimal element and $\{1, 2, ..., n\}^L$ as the maximal element. We will now consider the analogue of measure theory on the probability space $\left[\mathcal{P}(\{1, 2, ..., n\})\right]^L$.

[3]It satisfies the property of reflexivity, symmetry and transitivity.
[4]This explains the acronym "poset" for partially ordered sets used in the literature on random sets.

It can be useful to visualize this partial order when $L = 1$ under the form of a diagram with $n + 1$ layers corresponding to the size, i.e. number of elements of s. Such a diagram is given in Figure 1 for $n = 3$.

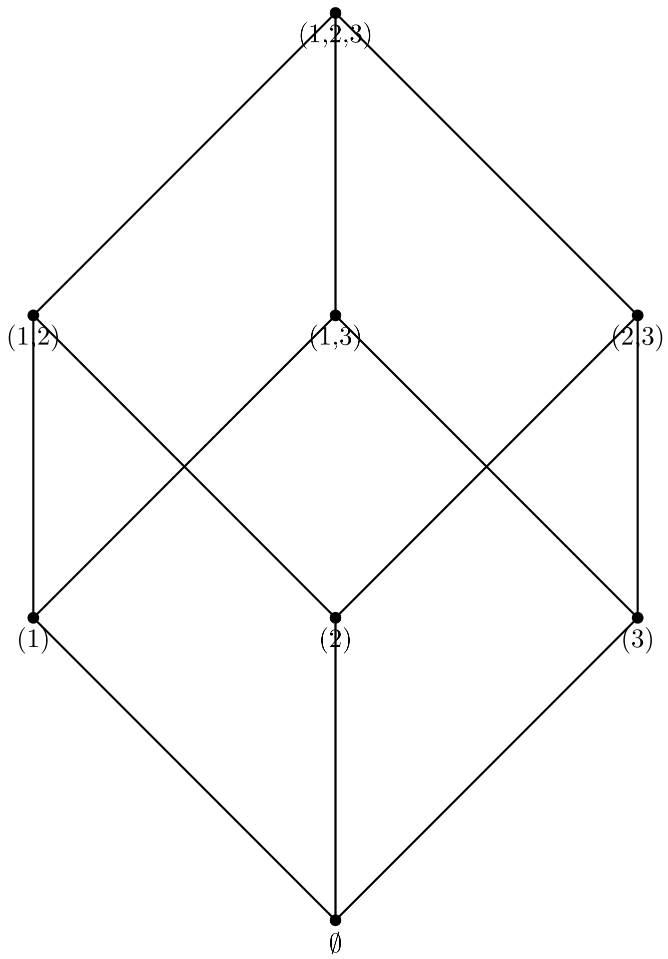

Figure 1: The ordering diagram, $L = 1, n = 3$.

If $L = 1$ we have $n + 1$ layers, 2^n subsets, and $n!$ ordered paths to go from $\emptyset$ to $\{1, ..., n\}$. When L is larger than 1, the inclusion order can be written equivalently as $s^* \supset s$, or $s^*(l) \supset s(l)$, for all $l = 1, ..., L$.

Let us introduce a nonnegative function v of set s.

Definition 1. *i)* The right cumulated function of v is the function:

$$V^+(s) = \sum_{s^* \supset s} v(s^*), \qquad \forall s. \tag{2.1}$$

The $p-$th right cumulated function of v is defined recursively by:

$$V^{(p)+}(s) = \sum_{s^* \supset s} V^{(p-1)+}(s^*), \quad p \geq 2,$$

with $V^{(1)+}(s) = V^+(s)$.

ii) The left cumulated function of v is the function:

$$V^-(s) = \sum_{s^* \subset s} v(s^*), \qquad \forall s. \tag{2.2}$$

The $p-$th left cumulated function of v is defined recursively by:

$$V^{(p)-}(s) = \sum_{s^* \subset s} V^{(p-1)-}(s^*),$$

with $V^{(1)-}(s) = V^-(s)$.

Example 3. For instance, if $L = 1, n = 3$, we have:

$$\begin{aligned}
V^+(\{1\}) &= v(\{1\}) + v(\{1,2\}) + v(\{1,3\}) + v(\{1,2,3\}), \\
V^+(\{1,2\}) &= v(\{1,2\}) + v(\{1,2,3\}), \\
V^-(\{1\}) &= v(\{1\}) + v(\emptyset), \\
V^-(\{1,2\}) &= v(\{1,2\}) + v(\{1\}) + v(\{2\}) + v(\emptyset), \\
V^{(2)+}(\{1\}) &= V^+(\{1\}) + V^+(\{1,2\}) + V^+(\{1,3\}) + V^+(\{1,2,3\}) \\
&= v(\{1\}) + 2v(\{1,2\}) + 2v(\{1,3\}) + 4v(\{1,2,3\}), \\
V^{(2)-}(\{1,2\}) &= V^-(\{1,2\}) + V^-(\{1\}) + V^-(\{2\}) + V^-(\emptyset) \\
&= v(\{1,2\}) + 2v(\{2\}) + 2v(\{1\}) + 4v(\emptyset).
\end{aligned}$$

Example 4. Let us assume that $v(s) = 1$. If $L = 1$, we have:

$$V^-(s) = 2^{n(s)}, V^+(s) = 2^{n-n(s)},$$

where $n(s)$ is the number of elements of s. These formulas are also valid for more variables. If $s = [s(1), ..., s(L)]$, we have:

$$V^-(s) = 2^{N(s)}, V^+(s) = 2^{N-N(s)},$$

where $N(s) = \sum_{l=1}^{L} n[s(l)], N = nL$.

The following property is proved in Appendix A.1.

Proposition 1. *It is equivalent to know the function v, or the function V^+ or V^-, or any other function $V^{(p)+}$ or $V^{(p)-}$.*

Proof. See Appendix A.1. □

There exist inversion formulas allowing to derive under closed form the function v from either V^+ or V^-, known as Moebius inversion formula [see Molchanov (2017), Theorem 1.1.61, Molchanov and Molinari (2018), p23]. For instance, if $L = 1$, we have:

$$v(s) = \sum_{s^* \subset s} \Big[(-1)^{n(s)-n(s^*)} V^-(s^*)\Big]. \tag{2.3}$$

However the recursive approach used in Appendix A.1 is more algorithmic and appropriate in practice to find the values of $v(s)$.

In particular, if $L = 1$ and v is a probability mass function on $\mathcal{P}[\{1, ..., n\}]$, then the function V^+ (resp. V^-) is the analogue of a survival function (resp. cumulative distribution function) and function v its "derivative". It is a decreasing (resp. increasing) function of s, in the sense that, if two subsets s_1, s_2 are ordered, $s_1 \subset s_2$, then:

$$V^+(s_1) \geq V^+(s_2), \qquad V^-(s_1) \leq V^-(s_2).$$

From now on we call $S-$decreasing function such a V^+ function and $S-$increasing function such a V^- function (S for set).

These interpretations can be extended to any order p. For instance $V^{(2)+}$ is a $S-$decreasing convex function and $V^{(2)-}$ is a $S-$increasing convex function, since, as mentioned after Proposition 3, v can be interpreted as an extension of the notion of derivative.[5] This $S-$notion of convexity does not assume any convexity of the probability space $\Big[\mathcal{P}(\{1, 2, ..., n\})\Big]^L$, or any mixture of set distributions [as in Fishburn (1974)].

Let us consider the total mass $\bar{V} = \sum_s v(s)$. We have:

Proposition 2. *For any subset s:*

$$V^+(s) + V^-(s) - v(s) \leq \bar{V}.$$

This inequality is a consequence of $v(s)$ being included in both definitions of V^+ and V^-, and of the partial order, since the summation in the definition of $\bar{V}$ includes also the $v(s^*)$ when

[5] For multivariate quantitative variables, these functions are said to be $p-$th order concave (resp. convex) [Denuit et al. (2013)].

the sets s^* and s cannot be compared (that is, when none of them is a subset of the other one).

Proposition 3. *Let us consider two nonnegative functions v, w and their cumulated counterparts V^+, V^- and W^+, W^-. We have:*

$$\sum_s v(s)W^+(s) = \sum_s V^-(s)w(s), \tag{2.4}$$

Similarly, by exchanging the roles of v and w, we also have: $\sum_s w(s)V^+(s) = \sum_s W^-(s)v(s)$.

Proof. See Appendix A.2. □

Equality (2.4) is an extension of the Abel Lemma [see Bromwich (2005), Chapter 20] and the analogue of the integration by part formula. For such an interpretation, let us consider two distributions on $\mathbb{R}^+$ with densities $f(x), f^*(x)$, c.d.f.'s $F(x), F^*(x)$, survival functions $S(x), S^*(x)$. We have:

$$\int_0^\infty F(x)f^*(x)\mathrm{d}x = -\int_0^\infty F(x)\mathrm{d}S^*(x) = -F(x)S^*(x)\Big]_0^\infty + \int_0^\infty f(x)S^*(x)\mathrm{d}x.$$

Since $-F(x)S^*(x)\Big]_0^\infty = 0$, we get: $\int_0^\infty F(x)f^*(x)\mathrm{d}x = \int_0^\infty f(x)S^*(x)\mathrm{d}x$. Equation (2.4) is the analogue of this equality. Thus, as mentioned earlier, the underlying function v is a kind of derivative (resp. opposite derivative) of function V^- (resp. V^+). With this interpretation, a function $W(s)$ that can be written as $W(s) = V^{(p)-}(s)$ is a function whose "derivatives" are nonnegative up to order p.

2.3 Distribution of a random set

When the variables $X = (X_{li})$ are stochastic, the associated set S is stochastic too. Then it is equivalent to define the distribution of X, or the distribution of S. We deduce from Proposition 1 that this distribution can be defined equivalently,

- by the elementary probabilities: $p(s)$, that are nonnegative and sum up to one across the different subsets s.
- by the associated $S-$decreasing (survival) function:

$$G^+(s) = \mathbb{P}[s \subset S] = \sum_{s^* \supset s} p(s^*), \quad \forall s, \tag{2.5}$$

- or by the associated $S-$increasing function (cdf):

$$G^-(s) = \mathbb{P}[S \subset s] = \sum_{s^* \subset s} p(s^*), \quad \forall s. \tag{2.6}$$

Then, for any nonnegative function v, with its associated $S-$increasing function V^- and $S-$decreasing function V^+, we deduce from Proposition 3 that:

$$\mathbb{E}[V^-(S)] = \sum_s p(s)V^-(s) = \sum_s v(s)G^+(s), \tag{2.7}$$

$$\mathbb{E}[V^+(S)] = \sum_s p(s)V^+(s) = \sum_s v(s)G^-(s). \tag{2.8}$$

When $L = 1$, the summation in eq. (2.5) is done on $2^{n-n(s)}$ terms, while it is done on $2^{n(s)}$ terms in eq. (2.6).

For instance, if we take $v(s) = p(s)$, we get:

$$\mathbb{E}[G^+(S)] = \mathbb{E}[G^-(S)] = \sum_s p(s)G^-(s) = \sum_s p(s)G^+(s) = G^{(2)+}(\emptyset) = G^{(2)-}(\{1, ..., n\}).$$

Then we can also interpret $\mathbb{E}[G^+(S)]$ by:

$$\mathbb{E}[G^+(S)] = \sum_s p(s)G^+(s) = \sum_s p(s) \sum_{s^* \supset s} p(s^*) = \mathbb{P}[S^* \supset S],$$

where S, S^* are i.i.d., with identical distribution $(p(s))$. In particular, since $G^+(s) + G^-(s) \leq 1$ for any s, we get: $G^{(2)+}(\emptyset) = \mathbb{P}[S^* \supset S] = G^{(2)-}(\{1, ..., n\}) \leq 1/2$. This is to be compared with continuous variables. Indeed, if Y, Y^* are i.i.d. with absolutely continuous distributions, then $\mathbb{P}[Y < Y^*] = 1/2$ by symmetry.

2.4 Law of Determinantal Point Process (LDPP)

Let us assume $L = 1$. An example of parametric family of distributions for random set S is the LDPP family. Its construction is based on the following Lemma:

Lemma 1 (Kulesza and Taskar (2012), Th. 2.1, Rising (2013), Theorem 2.3.1)**.** Let us consider a (n, n) matrix A and denote A_s the submatrix of A of dimension $\big(n(s), n(s)\big)$ including the rows and columns with indices in s. Then:

$$\det(Id + A) = \sum_s \det A_s. \tag{2.9}$$

More generally:

$$\det(I_{\bar{s}} + A) = \sum_{s^* \supset s} \det A_{s^*},$$

where $I_{\bar{s}}$ is the diagonal matrix with its $i-$th diagonal element equal to 1, if $i \in \bar{s}$, equal to 0, otherwise, where $\bar{s} = \{1, ..., n\} - s$ is the complement of s.

Let us now consider a (n, n) symmetric positive semi-definite matrix Σ. Any submatrix Σ_s of Σ is also positive semi-definite. Therefore all determinants $\det \Sigma_s$, i.e. principal minors of Σ, are nonnegative.

Definition 2. The LDPP family has the elementary probabilities:

$$p(s, \Sigma) = \frac{\det \Sigma_s}{\det(Id + \Sigma)}, \tag{2.10}$$

with the convention $\det \Sigma_\emptyset = 1$.

This family of $2^n - 1$ independent elementary probabilities is parametrized by Σ, that is by $\frac{n(n+1)}{2}$ independent parameters.

The associated $S-$decreasing (survival) function of the LDPP has a similar expression.

Proposition 4 (Borodin and Rains (2005), Brunel (2018))**.** *In the LDPP, we have:*

$$G^+(s) = \mathbb{P}[S \supset s] = \det K_s^+, \tag{2.11}$$

where the matrix K^+, often called kernel, is defined by:

$$K^+ = \Sigma(Id + \Sigma)^{-1} = Id - (Id + \Sigma)^{-1}. \tag{2.12}$$

If Σ is positive definite, the matrix K^+ is also symmetric positive definite and has all its eigenvalues strictly smaller than 1. It is in a one-to-one relationship[6] with Σ. Indeed, it is easily seen how to find Σ from K^+. We have:

$$\Sigma = -Id + (Id - K^+)^{-1} = K^+(Id - K^+)^{-1}. \tag{2.13}$$

Proposition 5. *In the LDPP family, the $p-$th right cumulated function is:*

$$G^{(p)+}(s) = \det\Big((p-1)I_{\bar{s}} + K^+\Big), \qquad \forall s, p \geq 2, \tag{2.14}$$

[6] The link between Σ and K is a bit more complicated if K has some eigenvalues equal to 1, see Online Appendix 2.4.

where $I_{\bar{s}}$ is the diagonal matrix with diagonal elements equal to 1, if $i \in \bar{s}$, and equal to 0, otherwise, and $G^{(1)+}(s) = G^+(s)$.

Proof. This is easily shown by induction using Lemma 1. For instance, for $p = 2$, we have:

$$\begin{aligned} G^{(2)+}(s) &= \sum_{s^* \supset s} G^+(s^*) \\ &= \sum_{s^* \supset s} \det K^+_{s^*} \\ &= \det[I_{\bar{s}} + K^+], \end{aligned}$$

by the extension of Lemma 1. □

Similar results can be obtained for the left cumulated functions $G^{(p)-}$ with another matrix $K^- = Id - K^+$ (see Appendix B.1). These functions measure recursively the integrated "left and right tails" of the set distribution.

Example 5 (Diagonal LDPP)**.** The matrix Σ is diagonal if and only if the matrix K^+ is diagonal. In this case, we have:

$$\begin{aligned} p(s) &= \prod_{i \in s} \sigma_{ii} / \prod_{i=1}^{n} (1 + \sigma_{ii}) = \prod_{i \in s} \frac{\sigma_{ii}}{1 + \sigma_{ii}} \prod_{i \in \bar{s}} [1 - \frac{\sigma_{ii}}{1 + \sigma_{ii}}], \\ G^+(s) &= \prod_{i \in s} \sigma_{ii} / \prod_{i \in s} (1 + \sigma_{ii}) = \prod_{i \in s} \frac{\sigma_{ii}}{1 + \sigma_{ii}}. \end{aligned}$$

One important property of LDPP as a model for set-valued variables is that it implies non-positive correlation between the component binary variables. Indeed, we have:

$$\begin{aligned} \text{Cov}[X_1, X_2] &= \mathbb{E}[X_1 X_2] - \mathbb{E}[X_1]\mathbb{E}[X_2] \\ &= G^+(\{1,2\}) - G^+(\{1\})G^+(\{2\}) = -(K^+_{12})^2 \leq 0. \end{aligned} \tag{2.15}$$

This is the so-called repulsive or diversity feature of the LDPP. For instance, if the indices 1 and 2 denote two firms and the binary variable measures whether or not they are in financial distress, then (2.15) can arise if the two firms are the main ones in a given industrial sector. Indeed, firm 1 can benefit from the bad results of firm 2, and in particular reinforce its monopolistic power.

3 Stochastic dominance and basket derivatives

3.1 Stochastic dominance

The standard notions of stochastic dominance can be extended to random sets and then indirectly to individual multivariate binary variables. In the standard framework of univariate, continuously valued risks, the stochastic dominance at order 1 is equivalently defined by comparing either the cumulative distribution function, or the survival function. In the set-valued framework, they no longer provide the same preference ordering. For expository purpose, we focus below on the notions based on the right cumulated functions $G^{(p)+}$ (see Appendix B.2 for the notion of left dominances based on left cumulated functions $G^{(p)-}$).

Definition 3. Let us consider two random sets S and S^*.

i) S^* is riskier than (i.e. right dominates) S at order 1, denoted $S^* \gtrsim_1 S$, if and only if $G^{*+}(s) \geq G^+(s), \forall s$.

ii) S^* is riskier than (i.e. right dominates) S at order 2, denoted $S^* \gtrsim_2 S$, if and only if $G^{*(2)+}(s) \geq G^{(2)+}(s), \forall s$.

These definitions can be equivalently defined by taking the complementary sets, since if S^* right dominates S at order 1 (resp. 2), then $\bar{S}$ right dominates $\bar{S^*}$. Thus, in the following, we are only interested in the case where for each index $i = 1, ..., n$, $i \in S$ means that individual i is a high risk.

Let us now consider a $S-$increasing function $V^-(s) = \sum_{s^* \subset s} v(s^*)$. By the summation by part (2.4) and eq. (2.7), we have:

$$\mathbb{E}[V^-(S)] = \sum_s v(s) G^+(s).$$

Therefore, if $S^* \gtrsim_1 S$, we have: $\mathbb{E}[V^-(S^*)] \geq \mathbb{E}[V^-(S)]$.

We deduce easily the following Proposition.

Proposition 6. *S^* is riskier than (i.e. right dominates) S at order 1, if and only if $\mathbb{E}[V^-(S^*)] \geq \mathbb{E}[V^-(S)]$ for any $S-$increasing function V^-.*

A similar result can be derived for the stochastic dominance at order 2.

Proposition 7. *S^* is riskier than S at order 2, if and only if $\mathbb{E}[V^{(2)-}(S^*)] \geq \mathbb{E}[V^{(2)-}(S)]$ for any $S-$increasing convex function $V^{(2)-}$.*

Proof. The result is obtained by applying twice the summation by part. Indeed we have:

$$\begin{aligned}\sum_s v(s)G^{(2)+}(s) &= \sum_s V^-(s)G^+(s)\\ &= \sum_s V^{(2)-}(s)p(s)\\ &= \mathbb{E}[V^{(2)-}(S)].\end{aligned}$$

The result follows. □

Since the set of $S-$increasing functions includes the set of $S-$increasing convex functions, we deduce the following corollary:

Corollary 1. *The right stochastic dominance at order 1 implies the right stochastic dominance at order 2.*

As already mentioned, it is possible to define left stochastic dominance. The right and left stochastic dominances define different partial orders (see Appendix B.2.2).

It is also possible to characterize a stochastic dominance by conditions on "virtual" joint distributions of the pair (S, S^*). Let us consider a pair of random sets S, S^* on $\mathcal{P}\big(\{1, ..., n\}\big)$. Then the joint distribution of (S, S^*) can be decomposed into the product of its marginal and conditional distributions as:

$$p(s, s^*) = p(s)p(s^*|s), \qquad \forall s, s^*.$$

Then we get a sufficient condition in terms of the conditional distribution of S^* given S to get S^* stochastically dominating S:

Corollary 2. *Let us assume:*

$$\mathbb{E}[V^-(S^*)|S] \geq V^-(S), \qquad \textit{for any S-increasing function } V^-, \tag{3.1}$$

then S^* dominates S at order 1.

Proof. This is a direct consequence of the iterated expectation theorem. We have:

$$\mathbb{E}[V^-(S^*)] = \mathbb{E}\Big[\mathbb{E}[V^-(S^*)|S]\Big] \geq \mathbb{E}\Big[V^-(S)\Big],$$

and by Assumption (3.1), the result follows. □

Similar results can be derived for the stochastic dominance at order 2, replacing the assumption (3.1) in Corollary 2 by the assumption at order 2:

$$\mathbb{E}[V^{(2)-}(S^*)|S] \geq V^{(2)-}(S), \qquad \text{for any S-increasing convex function } V^{(2)-}, \tag{3.2}$$

which replaces the standard convexity (Jensen-type) inequality. The proof is the same as the proof of Corollary 2.

The condition (3.1) in Corollary 2 means that the stochastic set S^* with conditional distribution $p(s^*|s)$ stochastically dominates at order 1 the constant subset $S = s$, for any s. Corollary 2 explains how to increase the risk by introducing a noisy observation S^* of set S.

Equivalent conditions can be written in terms of joint distribution of (S, S^*). More precisely,

Proposition 8 (see Strassen (1965), Th. 11, Kamae et al. (1977), Th. 1, Molchanov (2017), Th. 1.5.28(2))**.** *Let us denote by p_{S^*} and p_S the probability mass functions (pmf's) of S^* and S, respectively, then S^* dominates S at order 1, if and only if there exists a pair of random sets (S^*, S), such that the support of their joint distribution is in $\{(s, s^*) : s^* \supset s\}$ only, with the two marginal pmf's equal to p_{S^*} and p_S, respectively.*

Proposition 8 means that there exists a virtual probability space, on which S^* contains S almost surely, or equivalently $S^* = S \cup U$, where U is drawn in $\bar{S}$, and is interpreted as a "positive set noise".

Then the condition of stochastic dominance at order 2 can also be characterized in terms of virtual joint distribution.

Proposition 9. *S^* dominates S at order 2, if and only if there exists a virtual probability space for (S^*, S) such that the condition* (3.2) *is satisfied.*

This means that the additional noise U can be chosen such that:

$$\mathbb{E}[V^{(2)-}(S \cup U)|S = s] \geq V^{(2)-}(s), \qquad \text{for any } s, V^{(2)-}.$$

This corresponds to the mean-preserving spread in Rothschild and Stiglitz (1970) in a case where the expectation of S^* has no meaning[7], or to the Markov kernel property in Mosler and Scarsini (1991).

Let us recall that for real-valued absolutely continuous distributions, the stochastic dominance can be defined equivalently through either the cdf or the survival function. This is due to the

[7] The literature on random sets has introduced a notion of expectation of S, called Aumann or selection expectation [see Aumann (1965), Molchanov and Molinari (2018), Section 3]. This definition is not useful for our applications. Note also that for any subset s_1 of all individuals except one, say the first one, we have: $V^{(2)-}(\{1\} \cup s_1) - V^{(2)-}(s_1) = \sum_{s^* \subset s_1} V^-(\{1\} \cup s^*)$, which is increasing in s_1. This is a condition of increasing slope, called supermodularity in Perez (2018).

fact that the cdf and survival functions sum up to 1. Similarly, for set-valued variables, the stochastic dominance at orders 1 and 2 can also be defined from the left cumulated functions. However, in the set framework, the left and right notions of stochastic dominances are no longer equivalent, since the ordering of sets, or equivalently the ordering of multivariate binary variables, is a partial order. As a consequence, functions $G^+(\cdot), G^-(\cdot)$ [or equivalently the multivariate cdf and survival functions] no longer sum up to 1. We refer to Appendix B.2 for a detailed definition of left stochastic dominances and counterexamples of pairs such that one random set left (resp. right) dominates the other one, but not the other way round.

We also have the following proposition, linking the stochastic dominances between the set-valued variables and the stochastic dominances of their sizes.

Proposition 10. *If the set-valued variable S^* right dominates S at order 1, then the real-valued variable $n(S^*)$ stochastically dominates $n(S)$ at order 1 in the usual sense.*

This is an example of preservation of stochastic dominance under transformation [see Mosler and Scarsini (1991), Section 5].

Proof. Let us consider a function v which depends on set s through the size $n(s)$ only:

$$v(s) = g[n(s)], \forall s \subset \{1, 2, .., n\}. \tag{3.3}$$

Then:

$$V^-(s^*) = \sum_{s \subset s^*} v(s) = \sum_{k=0}^{n(s^*)} \binom{n}{k} g(k) := \sum_{k=0}^{n(s^*)} w(k) := W^-(n(s^*)), \tag{3.4}$$

where $w(k) = \binom{n}{k} g(k) \geq 0, \forall k$.

If S^* right dominates S at order 1, then for any function v satisfying condition (3.3), we have $\mathbb{E}[V^-(S^*)] \geq \mathbb{E}[V^-(S)]$. By equation (3.4), this implies that $\mathbb{E}[W^-(n(S^*))] \geq \mathbb{E}[W^-(n(S))]$, for any function W^- that is increasing on $\{1, ..., n\}$. In other words, $n(S^*)$ stochastically dominates $n(S)$ at order 1 in the usual sense. □

This result can be extended to the stochastic dominance at order 2.

3.2 Stochastic dominance in the LDPP framework

Let us consider two random sets S, S^* in the LDPP framework corresponding to kernels K^+ and K^{*+}, respectively. We deduce from Section 2.4 the characterization of stochastic dominance in the LDPP framework.

Proposition 11. *In the LDPP framework:*

i) S^* *(right) dominates* S *at order 1, if and only if,* $\det(K_s^{*+}) \geq \det(K_s^+), \quad \forall s$.

ii) S^* *(right) dominates* S *at order 2, if and only if,* $\det[I_{\bar{s}} + K^{*+}] \geq \det[I_{\bar{s}} + K^+], \quad \forall s$.

The inequalities in Proposition 11 are defining associated ordering on symmetric positive semi-definite matrices.

Definition 4. Let us consider two (n, n) symmetric positive semi-definite matrices Σ and Σ^*.

i) Σ^* determinantal dominates Σ at order 1, denoted $\Sigma^* \gg_{D1} \Sigma$, if and only if,

$$\det(\Sigma_s^*) \geq \det(\Sigma_s), \quad \forall s.$$

ii) Σ^* determinantal dominates Σ at order 2, denoted $\Sigma^* \gg_{D2} \Sigma$, if and only if,

$$\det[I_{\bar{s}} + \Sigma^*] \geq \det[I_{\bar{s}} + \Sigma], \quad \forall s.$$

Let us denote $\gg$ the usual (partial) ordering on symmetric matrices, i.e. the Loewner ordering.

Proposition 12. *i) For symmetric positive semi-definite matrices, the ordering* $\gg$ *implies* $\gg_{D1}$ *and* $\gg_{D2}$.

ii) For symmetric positive semi-definite matrices whose eigenvalues are all of absolute values smaller than 1, the order $\gg_{D1}$ *implies the order* $\gg_{D2}$.

Proof. i) $\Sigma^* \gg \Sigma$, if and only if $u'\Sigma^* u \geq u'\Sigma u, \quad \forall u \in \mathbb{R}^n$. By considering u_s, the vector with components of u for the indices in s, and zero components, otherwise, we get also $u_s'\Sigma^* u_s \geq u_s'\Sigma u_s, \quad \forall u_s$, or equivalently $\Sigma_s^* \gg \Sigma_s$. By the min-max theorem for symmetric matrices, this implies $\lambda_k^* \geq \lambda_k, k = 1, ..., n(s)$, where the λ_k (resp. λ_k^*) are the ranked eigenvalues of Σ_s (resp.Σ_s^*). In particular this implies: $\det \Sigma_s^* \geq \det \Sigma_s$.

Since $\Sigma^* \gg \Sigma$ is equivalent to $I_{\bar{s}} + \Sigma^* \gg I_{\bar{s}} + \Sigma$, the second implication is also obtained.

ii) This is a direct consequence of Corollary 1. □

We can also relate the set-valued stochastic dominance at order 1 of the LDPP with the usual stochastic dominance of the size of the LDPP (see Proposition 10). We first recall the following proposition concerning the size.

Proposition 13 (The law of the size of a LDPP, see Hough et al. (2006) and online Appendix 4)**.** *If* S *follows a LDPP model with kernel* K^+*, and* $\lambda_1^+ \geq \cdots \geq \lambda_n^+$ *are the eigenvalues of* K^+*, then the size* $n(S)$ *has the same distribution as the sum of* n *independent Bernoulli distributions with parameters* $\lambda_i^+, i = 1, ..., n$.

For two Bernoulli distributions with parameters p_1 and p_2, $\mathcal{B}(1, p_1)$ stochastically dominates at order 1 (in the usual stochastic dominance sense) $\mathcal{B}(1, p_2)$, if and only if $p_1 \geq p_2$. Moreover, the convolution of (real-)valued distributions preserves the stochastic dominance, in the sense that if X_1 (resp. Y_1) stochastically dominate X_2 (resp. Y_2), with X_1 (resp. X_2) being independent of Y_1 (resp. Y_2), then $X_1 + Y_1$ stochastically dominates $X_2 + Y_2$ at order 1. As a consequence, we have the following corollary, which is a direct consequence of Propositions 10 and 13:

Corollary 3. *If S^* and S both follow LDPP, with $n(S^*)$ and $n(S)$ following the convolution of Bernoulli variables with parameters $\lambda_1^{*+} \geq \cdots \lambda_n^{*+}$ and $\lambda_1^+ \geq \cdots \lambda_n^+$, respectively, then if*

$$\lambda_j^{*+} \geq \lambda_j^+, \qquad \forall j = 1, ..., n, \tag{3.5}$$

that is if the kernels of the two LDPP are such that K^{*+} is spectrally larger than K^+, or equivalently if Σ^* is spectrally larger than Σ, then $n(S^*)$ stochastically dominates $n(S)$ at order 1 in the usual sense.

Thus, in the LDPP framework both the right stochastic dominance and the spectral order between two LDPP's imply the usual stochastic dominance between the sizes of the LDPP (by Proposition 10 and Corollary 3, respectively). The first compares the principal minors of the two symmetric positive definite matrices, whereas the latter compares the eigenvalues of the two matrices.

The converse of Corollary 3 is not true. Indeed, the stochastic dominance of the convolution of $\mathcal{B}(1, \lambda_j^+), j = 1, ..., n$ over the convolution of $\mathcal{B}(1, \lambda_j^{*+}), j = 1, ..., n$ does not imply the inequalities (3.5).

The LDPP model and the associated stochastic dominances have interesting interpretations when the binary variables are close to independence. Let us write:

$$\Sigma = (\text{diag}\sigma)^{1/2}(Id + C)(\text{diag}\sigma)^{1/2},$$

where $\sigma = (\sigma_{11}, ..., \sigma_{nn})$ is the vector of diagonal elements of Σ, and matrix C has only zeros on the diagonal. Then we can perform a Taylor's expansion to approximate the set distribution.

Proposition 14. *Close to independence, we have:*

$$p(s, \Sigma) \approx \exp\left[c + \sum_{i \in s} \alpha_i + \sum_{i \in s} \sum_{j \in s, j < i} \beta_{i,j}\right],$$

with $\alpha_i = \log \sigma_{ii}$, $\beta_{i,j} = -c_{i,j}^2$ and c is a constant defined by the unit mass restriction.

Proof. We have:

$$\log p(s) = \text{constant} + \big[\sum_{i\in s} \log \sigma_{ii}\big] + \log\det(Id_s + C_s)$$
$$= \text{constant} + \big[\sum_{i\in s} \log \sigma_{ii}\big] - \frac{1}{2} Tr(C_s^2) + o(C^2)$$
$$= \text{constant} + \big[\sum_{i\in s} \log \sigma_{ii}\big] - \sum_{i\in s} \sum_{j\in s, j<i} c_{i,j}^2 + o(C^2).$$

The result follows.

□

Close to independence, the LDPP model is equivalent to a log-linear probability model with marginal effect α_i and (negative) pairwise interaction effects. Then the stochastic dominance can be written in terms of these effects.

3.3 Set derivatives

The $(X_{li}, l = 1, ..., L, i = 1, ..., n)$ can be interpreted as a portfolio of individual risks, usually called a basket. This basket, i.e. the associated random set, can be analyzed under the physical (historical) probability characterized by $p(s)$, $G^+(s)$, $G^-(s)$, $G^{(2)+}(s)$, $G^{(2)-}(s)$, used in insurance for fixing the premium, or under a risk-neutral probability characterized by $q(s)$, $Q^+(s)$, $Q^-(s)$, $Q^{(2)+}(s)$, $Q^{(2)-}(s)$, say, used in finance for pricing[8]. As usual, we can define different set (i.e. basket) derivatives. They differ by the basket design, since the basket can include either a small number of individuals (corporates), or a large number in an homogeneous segment (industrial sector, rating), and by the selected risks, such as solvency and/or liquidity risks). They also differ by the form of the payoff written on the associated random set.

3.3.1 Equivalent conditions for stochastic dominances

Let us first recall the "canonical derivatives" introduced on the derivative markets for real-valued risk variables. For two such variables Y_1 and Y_2 , the usual first-order stochastic dominance of Y_1 over Y_2 is equivalent to $\mathbb{E}[\phi(Y_1)] \geq \mathbb{E}[\phi(Y_2)]$ for any nondecreasing function ϕ [see Shaked and Shanthikumar (2007)]. But the set of all nondecreasing functions is the convex cone generated by the nondecreasing step functions of the form $\phi(y) = \mathbb{1}_{y>a}$, a varying. Thus we get a simpler, and equivalent characterization of the stochastic dominance at order 1:

$$\mathbb{E}[\mathbb{1}_{Y_1\geq a}] = \mathbb{P}[Y_1 \geq a] \geq \mathbb{P}[Y_2 \geq a] = \mathbb{E}[\mathbb{1}_{Y_2\geq a}], \qquad \forall a \in \mathbb{R}. \tag{3.6}$$

[8] We assume later on a zero risk-free rate. The results are easily extended to a nonzero risk-free rate.

This property motivates the introduction of tranches, that are digital derivatives with payoffs $\mathbb{1}_{Y_1 \geq a}$. Similarly, the usual stochastic dominance at order 2 of Y_1 over Y_2 is equivalent to

$$\mathbb{E}[(Y_1 - a)^+] \geq \mathbb{E}[(Y_2 - a)^+], \qquad \forall a \in \mathbb{R}, \tag{3.7}$$

which motivates the introduction of European calls (or stop-loss insurance policies).

Let us now discuss tranches and European calls in the set-valued framework. We first derive the following proposition, which provides the analogues of the equivalences (3.6) and (3.7), respectively.

Proposition 15. *i) The set of $S-$increasing functions V^- is the convex cone generated by the digit functions of the form $\mathbb{1}_{s_0 \subset s}$, with s_0 varying.*

ii) The set of $S-$increasing convex functions $V^{(2)-}$ is the convex cone generated by functions of the form

$$N(s_0, s) := Card(\{s^* : s_0 \subset s^* \subset s\}) = 2^{n(s)-n(s_0)} \mathbb{1}_{s_0 \subset s}, \tag{3.8}$$

with s_0 varying.

Proof. By Definition 1, we get:

$$\begin{aligned} V^-(s) &= \sum_{s_0 \subset s} v(s_0) = \sum_{s_0} v(s_0) \mathbb{1}_{s_0 \subset s}, \\ V^{(2)-}(s) &= \sum_{s^* \subset s} V^-(s^*) = \sum_{s^* \subset s} \sum_{s_0 \subset s^*} v(s_0) = \sum_{s_0} v(s_0) \Big[\sum_{s^*} \mathbb{1}_{s_0 \subset s^* \subset s} \Big] \\ &= \sum_{s_0} \Big[v(s_0) 2^{n(s)-n(s_0)} \mathbb{1}_{s_0 \subset s} \Big]. \end{aligned}$$

Thus the functions $\mathbb{1}_{s_0 \subset s}$, s_0 varying, generate the set of $S-$increasing functions, and the functions $2^{n(s)-n(s_0)} \mathbb{1}_{s_0 \subset s}$, s_0 varying, generate the set of $S-$increasing convex functions. □

Thus, S^* right dominates S at order 1, if and only if:

$$\mathbb{E}[\mathbb{1}_{s_0 \subset S^*}] \geq \mathbb{E}[\mathbb{1}_{s_0 \subset S}], \qquad \forall s_0. \tag{3.9}$$

This is the analogue of (3.6). Similarly, S^* right dominates S at order 2, if and only if

$$\mathbb{E}[2^{n(S^*)-n(s_0)} \mathbb{1}_{s_0 \subset S^*}] \geq \mathbb{E}[2^{n(S)-n(s_0)} \mathbb{1}_{s_0 \subset S}] \qquad \forall s_0,$$

which is the analogue of (3.7).

These equivalent conditions suggest the following definitions of tranches and European call options for set-valued risks.

3.3.2 Tranches

Definition 5. Let us fix two ordered sets $s_1 \subset s_2$. The payoff of the tranche (s_1, s_2) is 1 \$, if s is between (inclusively) s_1 and s_2, and 0 \$, otherwise.

The historical expected payoff of this tranche is: $G^+(s_1) - G^+(s_2) + p(s_2)$. Its price assuming a zero riskfree rate is: $Q^+(s_1) - Q^+(s_2) + q(s_2)$.

Let us consider $L = 1$ and the semi-upper tranche, that is a tranche with $s_2 = \{1, ..., n\}$. For instance, if $s_1 = \{1\}$, the tranche is paying 1\$, if $X_1 = 1$, and 0, otherwise. When the high risk is the speculative grade and the low risk the investment grade, this tranche is a type of Credit Default Swap (CDS) written on firm 1. If $s_1 = \{1, 2\}$, we get a CDS written on two firms and receive 1\$ if both 1 and 2 have speculative grades. Moreover, by (3.9), we deduce immediately the following corollary:

Corollary 4. *S^* dominates S at order 1 if and only if the expected payoffs of the tranche $(s_1, s_2 = \{1, ..., n\})$ is higher for S^* than for S, for any set s_1.*

When $s_2 = \{1, ..., n\}$, we have $Q^+(s_2) = q(s_2)$, and the price of such a semi-upper tranche is simply $Q^+(s_1)$. In particular, if S follows a LDPP under the risk-neutral measure, then we have the risk-neutral analogue of inequality (2.15):

$$Q^+(1, 2) - Q^+(\{1\})Q^+(\{2\}) = -(K_{12}^{Q+})^2 \leq 0,$$

where K_{12}^{Q+} is the $(1, 2)$ element of the risk-neutral kernel K^{Q+}. In other words, in the LDPP family, the price of the pairwise CDS with $s_1 = \{1, 2\}$ is not larger than the product of the two CDS prices with $s_1 = \{1\}$ and $s_1 = \{2\}$, respectively.

Remark 1. In the case where we have time series observations of sets (S_t), and, if tranche derivatives are issued at time t, maturing at time $t + 1$, say, then a tranche with $s_1 = S_t$ can be regarded as an "at-the-money" tranche. It pays 1 \$ if and only if $S_{t+1} \supset S_t$.

Remark 2. If $L = 2$, it is also possible to define tranches defined on two types of risks, as solvency risk and liquidity risk, say. Then a given corporate i can be in four different states $S_i = (S_i(1), S_i(2))$. For instance, $S_i = (\{1\}, \{1\})$, if it has both high solvency and liquidity risks. In such a framework, a tranche can concern both risks, i.e. the binary variable $X_i = X_{1,i}X_{2,i}$ and be written with one firm or two firms, for instance.[9]

[9] Since $X_{i,1}X_{i,2} = \min(X_{i,1}, X_{i,2})$, we can consider that this is a derivative written on the minimum of two

3.3.3 The European call option of the first kind

The following definition is motivated by Proposition 15.

Definition 6. The European call option of the first kind is the derivative with payoff $2^{n(S)-n(s_0)}\mathbb{1}_{s_0\subset S}$. In particular, the payoff is zero, if s_0 is not a subset of S.

This payoff function is the direct analogue of the stop-loss function for real-valued risks. Indeed, for a one-dimensional continuous risk U, the payoff of a European option with strike k is $(U-k)^+ = \max(U-k, 0)$. In the standard case the payoff is directional, measures a distance between U and k for $U > k$ and its price is obtained by considering the (risk-neutral) survival function at order 2:

$$Q^{(2)+}(k) = \int_k^\infty Q^+(u)\mathrm{d}u = \mathbb{E}^Q(U-k)^+. \tag{3.10}$$

In the set framework, the notions of difference and of positive part $(\cdot)^+$ do not exist. However, they have analogues. If $s \supset s_0$, the complement of s_0 in s is usually denoted $s - s_0$. However, even if $S - s_0$ is now well defined (when $s_0 \subset S$), the expectation of a random set has no meaning. In order to allow for expectation, we have to consider scalar transformation of $S - s_0$. The payoff $2^{n(S)-n(s_0)}\mathbb{1}_{s_0\subset S}$ can be equivalently written as $2^{n(S-s_0)}\mathbb{1}_{s_0\subset S}$, which is indeed a function of $S - s_0$.

This option has the following price formula, which is the analogue of (3.10):

Proposition 16. *The expected payoff (resp. price) of the European call with $S-$strike s_0 is $G^{(2)+}(s_0)$ [resp. $Q^{(2)+}(s_0)$].*

Proof. By using the same argument as in the proof of Proposition 15, we have:

$$\begin{aligned} G^{(2)+}(s_0) &= \sum_s \Big[p(s)\sum_{s^*}\mathbb{1}_{s\supset s^*\supset s_0}\Big] \\ &= \sum_s \Big[p(s)N(s_0,s)\Big] \\ &= \mathbb{E}[N(s_0,S)], \end{aligned}$$

where $N(s_0,S) = 2^{n(s)-n(s_0)}\mathbb{1}_{s_0\subset s}$ is the payoff of the European call. The result follows for the expected payoff. The proof is similar for the price under the risk-neutral probability. □

We deduce immediately the following corollary:

---

risks [see Molchanov and Schmutz (2010) for such derivatives written on two quantitative risks]. In our framework of $0-1$ variables, by considering all min-max options written on the X_i's, we generate all basket derivatives [see Molchanov and Schmutz (2014) for a similar result for quantitative risks].

Corollary 5. *S^* dominates S at order 2 if and only if the expected payoffs of the European call are larger for S^* than for S, for any $S-$strike s_0.*

3.3.4 Alternative definitions of European option

The previous definition of European calls is an analogue of the standard European calls for real-valued risks, from the point of view of stochastic dominance. However, it has some downsides.

The first downside of the European call of the first kind is that the power function of base 2 could be difficult to understand for market participants. Indeed, this implies that the payoff function can take large values, if $n(S)$ is large and $s_0 \subset S$. This issue can be easily addressed by considering other scale transformations of $S - s_0$, when $s_0 \subset S$. A first possible candidate is the size, which leads to the payoff $n(S - s_0)\mathbb{1}_{s_0 \subset S}$. A second one is to replace $2^{n(S-s_0)}$ by another exponential transform $\exp[u\, n(S - s_0)]$, where u is chosen to get a more reasonable value of the payoff. In particular, if $u = \ln 2$, we recover the initial definition. The motivation of using exponential function is that, in the real-valued framework, it is known that such exponential European options are equivalent to, and easier to compute than the standard call options [see the transform inversion formula of Duffie et al. (2000), Proposition 2].

A second issue is that the payoff is non zero only when $s_0 \subset S$. Thus the probability that the payment is triggered is very small, especially when $n(s_0)$ is large. An alternative condition of triggering a payment is $S \cap \bar{s_0} \neq \emptyset$. This new condition has two motivations. First, because $s_0 \subset S$ implies $S \cap \bar{s_0} \neq \emptyset$, the new trigger condition is weaker, and corresponds to a more appealing insurance or financial interpretation. Suppose that the n binary risks can be ranked in increasing importance and assume s_0 corresponds to the set of risks of lower importance, while $\bar{s_0}$ corresponds to those of higher importance. Then the condition $s_0 \subset S$ means that payment is triggered when all the low importance events, as well as one of the high importance events is realized, whereas $S \cap \bar{s_0} \neq \emptyset$ means that payment is triggered so long as at least one high importance events is realized.

Moreover, when S does not necessarily include s_0, the analogue of $S - s_0$ is $S \cap \bar{s}_0$, which motivates the trigger condition $S \cap \bar{s_0} \neq \emptyset$. In the following, we will denote the "hitting set" $S \cap \bar{s}_0$ by $(S - s_0)^+$, since it is equal to zero if and only if the new trigger condition is not satisfied.

According to the above discussions, several candidate payoff functions can be considered:

- $n(S \cap \bar{s}_0) = n\big[(S - s_0)^+\big]$;
- $\exp[u\, n(S \cap \bar{s}_0)] - 1$;
- $\exp[u\, n(S \cap \bar{s}_0)]\mathbb{1}_{S \cap \bar{s_0} \neq \emptyset}$.

Among these three payoff designs, the second and third payoffs are quite similar, while the first function can be obtained as a limiting case of the second one. Indeed, when u is small, we get:

$$\exp[u\, n(S \cap \bar{s}_0)] - 1 \approx u\, n\big[(S - s_0)^+\big],$$

and recover the first payoff design. As a consequence, we follow below the second approach.

Definition 7. The European call option of the second kind with strike s_0 and weight u is the basket derivative with payoff $\exp[u\, n(S \cap \bar{s}_0)] - 1 = \prod_{i \in S \cap \bar{s_0}} \exp(uX_i) - 1$.

These payoffs depend on the hitting set: $(S - s_0)^+ = S \cap \bar{s_0}$, represented by the binary variables $X_i, i \in S \cap \bar{s_0}$. These payoffs only depend on the $n(\bar{s_0})$-dimensional marginal distribution of the random set on $\bar{s_0}$. This marginalization is especially simple in the LDPP framework, since the hitting set follows also a LDPP with the kernel $K^+_{\bar{s_0}}$.

By limiting the payoffs to another family indexed by s_0, or equivalently to the closed positive cone generated by these functions, we implicitly exclude "pathological" functions not likely representing the preferences of the investors and that can lead to too restrictive stochastic dominances [see e.g. the notions of integral stochastic ordering in Marshall (1991), Müller (1997), Denuit and Mesfioui (2010) and of almost stochastic dominance in Tsetlin and Winkler (2018)].

3.3.5 Expected shortfall

We have also the decomposition of the expected payoff (resp. the price) of the option in terms of the expected occurrence (resp. the price of the semi-upper tranche on digital option) and the expected shortfall (resp. the cost of high risk). Indeed,

- for the European option of the first kind, we have:

$$G^{(2)+}(s_0) = G^+(s_0) ES_1(s_0),$$

  where $ES_1(s_0) = \mathbb{E}[N(s_0, S) | s_0 \subset S]$ is an expected shortfall.

- for the European option of the second kind, we have:

$$\mathbb{E}\big\{\exp[un(S \cap \bar{s_0})] - 1\big\} = \mathbb{P}[S \not\subseteq s_0] ES_2(s_0),$$

  where

$$ES_2(s_0) := \frac{\mathbb{E}\big\{\exp[un(S \cap \bar{s_0})] - 1\big\}}{\mathbb{P}[S \not\subseteq s_0]} = \mathbb{E}\big\{\exp[un(S \cap \bar{s_0})] - 1 \mid S \not\subseteq s_0\big\}.$$

is the expected shortfall.

Remark 3. In order to differentiate the n binary variables, it is also possible to extend the above payoff function by considering $\prod_{i \in s \cap \bar{s_0}} \exp(u_i X_i) - 1$, where $(u_1, u_2, ..., u_n)$ is a vector of fixed parameters. Its expected payoff is, up to the constant -1, the marginal Laplace transform of $X_i, i \in \bar{s_0}$ (see online Appendix 2 for the expression of the Laplace transform).

3.3.6 The order implied by the European call option of the second kind

Proposition 17. *S^* dominates S at order 1 if and only if for a sufficiently large weight u, the expected payoffs of the European call of the second kind for S^* are non smaller than for S, for any $S-$strike s_0.*

Proof. The expected payoff is the Laplace transform of the marginalized distribution of $(S - s_0)^+$. Thus by (eq. a.22) in Online Appendix 2 on Laplace transform, for large, positive u, the expected payoff of the European option for for S^* (resp. S) has the dominant term $\exp\Big\{u[n - n(s_0)]\Big\}\mathbb{P}[X_i^* = 1, \forall i \in \bar{s_0}]$ (resp. $\exp\Big\{u[n - n(s_0)]\Big\}\mathbb{P}[X = 1, \forall i \in \bar{s_0}]$). Thus:

$$\mathbb{P}[X_i^* = 1, \forall i \in \bar{s_0}] \geq \mathbb{P}[X = 1, \forall i \in \bar{s_0}],$$

or equivalently $G^{*+}(\bar{s_0}) \geq G^+(\bar{s_0})$. Thus by varying s_0, we deduce that S^* dominates S at order 1.

Conversely, if S^* dominates S at order 1, then, when we compare the expansion of the expected payoff for options written on S^* and S using (eq. a.22), the coefficient in front of each term for S^* dominates the corresponding coefficient for S. Thus the expected payoff of the European option for S^* is non smaller than the expected payoff of the European option for S^*. □

This result is to be compared with its analogue for the European options of the first kind, which induces an order equivalent to the left stochastic dominance at order 2 (see Corollary 5).

3.3.7 Put-call parity

For real-valued risks, because of the put-call parity, it is equivalent to know the prices of all the call options or the prices of all the put options. In our framework, instead of using the hitting set $(S - s_0)^+ = S \cap \bar{s_0}$, put options can be defined through the “missing set” $(s_0 - S)^+ = s_0 \cap \bar{S}$.

Let us for instance consider put and call options based on the cardinality of the sets $(S - s_0)^+ = S \cap \bar{s_0}$ and $(s_0 - S)^+ = s_0 \cap \bar{S}$. Since the set $(s_0 - S)^+ \cup (S - s_0)^+$ has no simple expression, we cannot expect a put-call parity with put and call with a same $S-$strike s_0. Nevertheless, analogues of put-call parity relation can be obtained by combining put and call with complementary $S-$strikes. For instance, we have the identity:

$$(s_0 - S)^+ \cup (S - \bar{s_0})^+ = s_0, \Rightarrow n[(s_0 - S)^+] + n[(S - \bar{s_0})^+] = n(s_0),$$

hence:

$$\mathbb{E}\big\{n[(s_0 - S)^+]\big\} + \mathbb{E}\big\{n[(S - \bar{s_0})^+]\big\} = n(s_0). \tag{3.11}$$

This equality can be understood as a put-call parity.

Similarly, we have another kind of put-call parity:

$$(S - s_0)^+ \cup (S - \bar{s_0})^+ = S, \Rightarrow n[(S - s_0)^+] + n[(S - \bar{s_0})^+] = n(S), \tag{3.12}$$

which links the distribution of the size $n(S)$ to the price of two call options with complementary $S-$strikes through:

$$\mathbb{E}\big\{n[(S - s_0)^+]\big\} + \mathbb{E}\big\{n[(S - \bar{s_0})^+]\big\} = \mathbb{E}[n(S)]. \tag{3.13}$$

3.4 European options in the LDPP framework

In this section we consider the price of the two kinds of European options introduced in Sections 3.3.3 and 3.3.4 under the LDPP assumption. Specifically, it is shown that both options allow for closed form expressions, thanks to the closure of the LDPP under marginalization and conditioning, respectively.

i). **European options of the first kind** Let us represent, for any random set $S \subset \{1, ..., n\}$ such that $S \supset s_0$, the subset of binary variables $X_i, i \in S \cap \bar{s_0}$ by another set-valued variable $T = S - s_0$. Then the expected shortfall component $\mathbb{E}[N(S, s_0)|S \supset s_0] = \mathbb{E}[\prod_{i \in S \cap \bar{s_0}} 2^{X_i}|S \supset s_0]$ is the probability generating function of variable T, conditional on $S \supset s_0$. This conditional distribution is still LDPP in the LDPP framework:

Proposition 18. *Conditional distribution of a LDPP [see Affandi et al. (2012), eq. (6)].*

If S follows a LDPP model with kernel K^+, then conditional on $S \supset s_0$, the variable $S - s_0$ follows another LDPP model on the set $\bar{s_0} = \{1, ..., n\} - s_0$, with kernel $K^+(s_0)$ given by:

$$K^+(s_0) = \Big[Id - (\Sigma + I_{\bar{s_0}})^{-1}\Big]_{\bar{s_0}}, \tag{3.14}$$

where $I_{\bar{s_0}}$ is the (n, n) diagonal matrix with ones on the entries indexed by $\bar{s_0}$ and zero otherwise.[10]

[10] The (n, n) diagonal matrix $I_{\bar{s_0}}$ differs from the identity matrix $Id_{\bar{s_0}}$. The first one has dimension (n, n), the

Thus it suffices to compute the pgf of a LDPP. This (closed form) expression is given in eq. (eq. a.27) in online Appendix 2 and we get:

$$\mathbb{E}[N(S,s_0)|S\supset s_0] = \det\left[Id_{\bar{s_0}} + K^+(s_0)\right].$$

Remark 4. The expression (3.14) of the conditioning kernel is greatly simplified if the individuals are ordered such that the set s_0 includes the first $n(s_0)$ individuals and $\bar{s}_0$ the last ones. If the associated block decomposition of K^+ is $K^+ = \begin{bmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{bmatrix}$, then $K^+(s_0) = K_{22} - K_{21}K_{11}^{-1}K_{12}$.

ii). **European options of the second kind** Because the expected payoff of these options is the marginal Laplace transform of $X_i, i \in \bar{s_0}$, we rely on the property that, if $(X_1, ..., X_n)$ follows a LDPP with K^+, then $X_i, i \in \bar{s_0}$ also follows a LDPP with the kernel $K^+_{\bar{s_0}}$. Thus by (eq. a.29), we get:

$$\mathbb{E}[\exp[u\,n(S\cap\bar{s_0})]-1] = \det\left[Id + K^+_{\bar{s_0}}(e^u-1)]\right] - 1.$$

As a summary, the main closed form prediction (resp. pricing) formula in the LDPP framework are gathered in Table 2 below. They are provided for prediction with kernel K. Their analogues for pricing are obtained by replacing K by $K^{\mathbb{Q}}$. We also provide the formulas when the risks $1, ..., n$ are weighted by $u := (u_1, u_2, ..., u_n)'$. These weights have a natural interpretation in the case of default risk, where these weights are the individual expected loss-given-default, or in the case of cyber risk, where the weight u_i is an expected loss given a cyber attack on the firm (or computer system) i.

Name	Payoff	Prediction
Tranche	$\mathbb{1}_{s_1\subset S\subset s_2}$	$\det K_{s_1} - \det K_{s_2} + \frac{\det\Sigma_{s_2}}{\det(Id+\Sigma)}$
Eur. option 1st kind	$2^{n(S)-n(s_0)}\mathbb{1}_{s_0\subset S}$	$\det K_{s_0} \det\left[Id_{\bar{s_0}} + K^+(s_0)\right]$
Size of hitting set	$n(S\cap\bar{s_0})$	$\mathrm{Tr}(K_{\bar{s_0}})$
Eur. option 2nd kind	$\exp\left[u\,n(S\cap\bar{s_0})\right]-1$	$\det\left[Id + K^+_{\bar{s_0}}(e^u-1)]\right] - 1$
(Weighted) exponential	$\exp(-u(S))$	$\det\left[Id + K^+(e^u-1)]\right]$

Table 2: Prediction formulas for the main derivatives in the LDPP framework

4 Exchangeability

Let us first consider the case $L = 1$ for expository purpose. It is usual to consider homogeneous groups of individuals (contracts). In mathematical terms, a group of individuals is homogeneous,

second one dimension $(n(\bar{s_0}), n(\bar{s_0}))$. They are related by: $Id_{\bar{s_0}} = \left[I_{\bar{s_0}}\right]_{\bar{s_0}}$.

if and only if the distribution of $(X_i, i = 1, ..., n)$, or equivalently of S, is exchangeable, that is invariant by permutation of the individual indices. Then the analysis can be extended to clusters (or blocks) of homogeneous groups.

4.1 The i.i.d. cross-sectional model

The simplest example of exchangeability is the i.i.d. cross-sectional model, where the binary variables X_i are independent with the same Bernoulli distribution $\mathcal{B}(1, \pi)$. Then we have:

$$p(s) = \pi^{n(s)}(1-\pi)^{n-n(s)},$$

where $n(s)$ is the number of elements in s, i.e. the size of s. This distribution depends on set s through the size $n(s)$, which is a function of s invariant by permutation.

The stochastic size $n(S)$ follows the binomial distribution distribution $\mathcal{B}(n, \pi)$. In this i.i.d. cross-sectional model, it is possible to replace the analysis of s by the analysis of its size, or equivalently the partial order on s by the total order on $n(s)$.

4.2 De Finetti's representation theorem

The exchangeable models are closely related to the i.i.d. cross-sectional models.

Proposition 19. *de Finetti's Representation Theorem [see Heath and Sudderth (1976)]*

For an infinite space $\{1, 2, ...\}$, the distribution of S is exchangeable if and only if S follows an i.i.d. cross-sectional model conditional on a stochastic intensity π.

When the state space $\{1, ..., n\}$ is finite, and the X_i's are cross-sectionally i.i.d. with stochastic intensity, the elementary probabilities can be written as:

$$p_n(s) = \int_0^1 \pi^{n(s)}(1-\pi)^{n-n(s)} g(\pi) \mathrm{d}\pi, \tag{4.1}$$

where $g(\cdot)$ denotes the density of the stochastic intensity.

As noted above the size is invariant by permutation and it is easily seen from the formula of the elementary probabilities that the distribution of S given $n(S) = n_0$ is uniform with elementary probabilities $1/\binom{n}{n_0}$. We deduce the following corollary:

Corollary 6. *Under stochastic intensity, the distribution of S is characterized by the distribution of the size:*

$$P(\widetilde{n}) = \mathbb{P}[n(S) = \widetilde{n}], \qquad \widetilde{n} = 0, ..., n.$$

Then $p(s) = P(n(s))/\binom{n}{n(s)}, \forall s.$

In the stochastic intensity model (4.1), we have:

$$\text{Cov}(X_i, X_j) = G^+(\{i,j\}) - G^+(\{i\})G^+(\{j\}) = \mathbb{E}[\pi^2] - (\mathbb{E}[\pi])^2 = \text{Var}(\pi) \geq 0.$$

Thus we get a "positive" dependence between the binary variables X_i and X_j for any pair i, j. For exchangeable model, this is the only possible scenario, when $n = \infty$, by de Finetti's theorem. However, negative dependence can exist in finite space with exchangeable model, as seen below with the LDPP model.

4.3 Exchangeability in the LDPP framework $(L = 1)$

In the LDPP framework and finite state space $\{1, ..., n\}$, the exchangeability condition can be written in terms of the kernel matrix.

Proposition 20. *The LDPP model is exchangeable if and only if* $K^+ = \alpha^+ Id + \beta^+ J$*, where* J *is the* (n, n) *matrix whose elements are all equal to 1, and the parameters* α *and* β *satisfy the constraint* $\alpha^+ + n\beta^+ \leq 1$.

The exchangeable matrix has $n - 1$ eigenvalues equal to α^+ and one eigenvalue equal to $\alpha^+ + n\beta^+ \leq 1$. It is positive semi-definite if $\alpha^+ \geq 0$ and $\alpha^+ + n\beta^+ \geq 0$.

Corollary 7. *In the exchangeable LDPP framework, the distribution of the size is the distribution of the sum of two independent variables following a binomial distribution* $\mathcal{B}(n - 1, \alpha^+)$ *and a Bernoulli distribution* $\mathcal{B}(1, \alpha^+ + n\beta^+)$*, respectively.*

Proof. This is a direct consequence of Propositions 13 and 20. □

4.4 Block models

Whereas the LDPP framework implies repulsive features, the stochastic intensity model implies positive dependence. These two features can be managed together for more flexibility in block models.

Let us consider the case of two blocks $k = 1, 2$, and construct a law of random set such that we get homogeneous blocks and the possibility of repulsion between blocks. This can be done by considering latent binary variables. For $L = 1$, we get:

$$X_i = Z_i Y_{1i} + (1 - Z_i) Y_{2i}, \qquad i = 1, ..., n, \tag{4.2}$$

where

- $Z_i = 1$, if i belongs to block 1, and $Z_i = 0$, otherwise,

- $Y_{k,i} = 1$, if i is high risk conditional on belonging to block k, and $Y_{k,i} = 0$, otherwise, for $k = 1, 2$.

Definition 8. The LDPP with two homogeneous blocks is the model (4.2), in which:

i) $Z = (Z_1, ..., Z_n)'$, $Y_k = (Y_{k,1}, ..., Y_{k,n})', k = 1, 2$ are independent.

ii) Z follows a LDPP with kernel K^+, $(Y_{k,1}, ..., Y_{k,n})$ follows a stochastic intensity model with stochastic intensity parameter, with distribution g_k, $k = 1, 2$ [see(4.1)].

In this block model, the binary variables Z_i, Y_{1i}, Y_{2i} are latent. The LDPP $Z = (Z_1, ..., Z_n)'$ partitions $\{1, ..., n\}$ into two latent blocks $\mathcal{P}_1$ and $\mathcal{P}_2$, where $i \in \mathcal{P}_1$, if $Z_i = 1$, and $i \in \mathcal{P}_2$, otherwise. The observed set S can be decomposed into two latent subsets $S = S(1) \cup S(2)$, where $S(k) = S \cap \mathcal{P}_k, k = 1, 2$, is the set of indices i belonging to $\mathcal{P}_k$ such that $Y_{k,i} = 1$. Note that $\mathcal{P}_k, S(k)$, as well as their sizes are random.

Let us now characterize the distribution of $n(S)$ through this latent representation.

Distribution of $N_1 := n(\mathcal{P}_1)$ and $N_2 := n(\mathcal{P}_2) = n - N_1$. Because Z follows a LDPP, N_1 is simply the size of this LDPP and by Proposition 13, its distribution is the convolution of n Bernoulli distributions. Let us denote by $\pi_1(n_1)$ its p.m.f.

Distribution of $n(S(1)), n(S(2))$ given $N_1 = n_1$. Because Y_1 and Y_2 are independent, the joint distribution of $n(S(1)), n(S(2))$ given $N_1 = n_1$ is:

$$
\begin{aligned}
&\mathbb{P}[n(S(1)) = \tilde{m}_1, n(S(2)) = \tilde{m}_2 | N_1 = n_1] \\
&= \mathbb{P}[n(S(1)) = \tilde{m}_1 | N_1 = n_1] \mathbb{P}[n(S(2)) = \tilde{m}_2 | N_2 = n - n_1] \\
&= \underbrace{\binom{n_1}{\tilde{m}_1} \Big[\int_0^1 \pi^{\tilde{m}_1} (1-\pi)^{n_1 - \tilde{m}_1} g_1(\pi) \mathrm{d}\pi \Big]}_{:= \pi_{1|1}(\tilde{m}_1 | n_1)} \underbrace{\binom{n - n_1}{\tilde{m}_2} \Big[\int_0^1 \pi^{\tilde{m}_2} (1-\pi)^{n - n_1 - \tilde{m}_2} g_2(\pi) \mathrm{d}\pi \Big]}_{:= \pi_{2|1}(\tilde{m}_2 | n - n_1)}.
\end{aligned}
$$

Distribution of $n(S)$ given $N_1 = n_1$. Because $n(S) = n(S(1)) + n(S(2))$, the conditional distribution of $n(S)$ is simply the convolution of the conditional distributions of $n(S(1))$ and $n(S(2))$:

$$
\mathbb{P}[n(S) = \tilde{n} | N_1 = n_1] = \sum_{\tilde{m}_1 = 0}^{\tilde{n}} \pi_{1|1}(\tilde{m}_1 | n_1) \pi_{2|1}(\tilde{n} - \tilde{m}_1 | n - n_1).
$$

Distribution of $n(S)$. Finally, the distribution of $n(S)$ is obtained by integrating out the above conditional distribution. We get:

$$\begin{aligned}\mathbb{P}[n(S)=\tilde{n}] &= \sum_{n_1=0}^{n} \pi_1(n_1)\mathbb{P}[n(S)=\tilde{n}|N_1=n_1] \\ &= \sum_{n_1=0}^{n} \pi_1(n_1) \sum_{\tilde{m_1}=0}^{\tilde{n}} \pi_{1|1}(\tilde{m_1}|n_1)\pi_{2|1}(\tilde{n}-\tilde{m_1}|n-n_1).\end{aligned}$$

5 Summary statistics and interpretations

In the applications, the dimensions n, T, and/or L can be very large and the data as well as the distribution of the random set difficult to visualize. The aim of this section is to discuss such visualizations, whose use and interpretation generally depend on the topic of interest. They can be adjacency plots of the data, distributions of size variables, set variance-covariance matrices, or derivative prices. In order to learn about their use, we provide them for different schemes of set distributions. These are:

1) The independence scheme, that can depend on n and on the parameter p of the Bernoulli distribution (Section 4.1).

2) The stochastic intensity scheme, that can depend on n and on the distribution of π (Section 4.2).

3) The block models to mix the positive dependence within blocks and the repulsive effect between blocks (Section 4.4).

4) A LDPP scheme in which the matrix Σ has a factor representation (see Appendix D).

These summaries are a first step before developing a coherent exploratory data analysis that would help in specifying the distribution of the set [Gouriéroux and Lu (2023b)].

5.1 Adjacency plots

The adjacency plots have been initially introduced to visualize networks. They can also be used to visualize panel data. We consider below this latter application.

5.1.1 Panel data

If $L=1$, the adjacency plots can be used for representing the values of X for time t (in the $x-$axis) and individual i (in the $y-$axis), that is to visualize the data. This allows to see the evolution $S_t, t=1,...,T$ of the random set. They are the analogues of the series of returns usually reported for quantitative risks, but require a specific analysis to account for the binary feature. To illustrate these point patterns, we consider the i.i.d. framework of Section 5 for the S_t's and provide simulated plots corresponding to the following schemes.

- Scheme 1: The $X_{i,t}$'s are i.i.d. $\mathcal{B}(1, \pi)$ distributed.
- Scheme 2: The S_t are independent with a LDPP distribution with factor representation, i.e. $\Sigma = \sigma^2 Id + \lambda\beta\beta'$, where $\beta = (1, ..., n)$, $\lambda = 5$, $\sigma = 0.1$. In other words, $\lambda\beta\beta'$'s contribution to Σ is significantly larger than that of $\sigma^2 Id$. Because $\lambda\beta\beta'$ is of rank 1 and for an LDPP whose matrix is of rank 1, the size is equal to 1 almost surely. Thus, we expect that for this specification of Σ, the size of the LDPP takes value 1 most of the time.

Figure 2 displays a plot for Scheme 1, and two plots for Scheme 2, where the individuals are not ranked ex-ante, and ranked ex-ante by means of the factor values, respectively. We also choose $\pi = 0.002$ such that the expected number of points is the same through the two schemes.

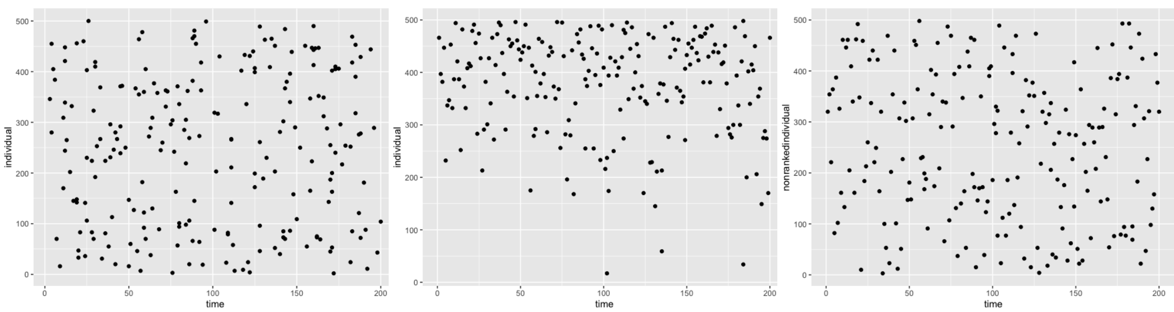

Figure 2: Left panel: the adjacency plot for an independent Bernoulli model. Middle panel: LDPP with one factor representation, where the individuals are ranked by means of the factor values. Right panel: the same LDPP, where the individuals are unranked. In other words, the middle and right panels correspond to the same LDPP up to a reordering of individuals.

We also display, for each of the three plots above, the two marginal distributions, that are the size of S_t (in the time series dimension), as well as the size of $S_i, i = 1, ..., n$ (in the cross-sectional dimension).

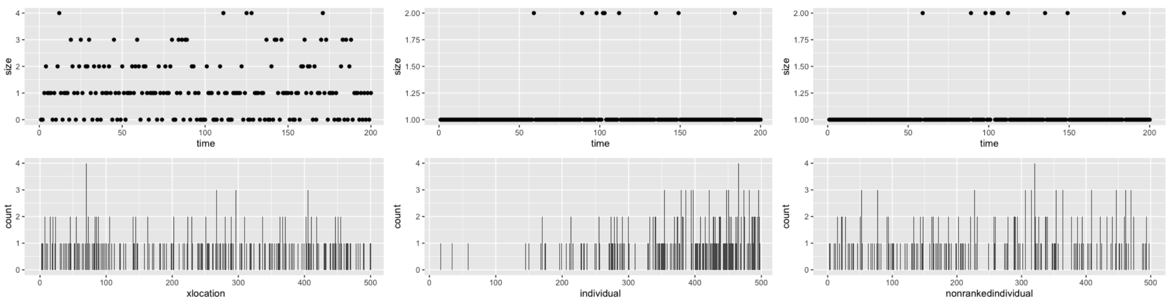

Figure 3: Size of $S_t, t = 1, ..., T$ and the size of $S(i), i = 1, ..., n$ for the three plots in Figure 3. Left panel: the independent Bernoulli model. Middle and right panel: the LDPP model with ranked and unranked individuals, respectively.

The middle and right panel have the same time series for $n(S_t)$, which is expected, since at each time t, the reordering of the individuals does not impact the size of $n(S_t)$. Moreover, this size takes value 1 most of the time, except on several occasions where we have $n(S_t) = 2$. In the left panel, the path of $n(S_t)$ is more erratic, with a larger range $\{0, ..., 4\}$.

In the time series dimension, these plots are not very informative on the underlying structure of cross-sectional dependence, since the observations are made very noisy due to the assumption of serial independence. They will be much more informative if there are both structures of cross-sectional and serial dependence.

The middle and right panel differ, however, in terms of the plot of $S(i)$. In the middle panel, we see an increasing tendency, since the elements of β are increasing. This is no longer the case in right panel, where the individuals are not ranked.

5.1.2 Joint adjacency plots

Such adjacency plots can also be used to analyze jointly two types of variables, i.e. with $L = 2$ and n individuals. In such a framework, the data are entirely visualized with four adjacency plots corresponding to the binary variables $Z_1 = X_1X_2$, $Z_2 = X_1(1 - X_2)$, $Z_3 = (1 - X_1)X_2$ and $Z_4 = (1 - X_1)(1 - X_2)$. Note, that these four binary variables are linked through:

$$Z_4 = (1 - Z_1 - Z_2)(1 - Z_1 - Z_3).$$

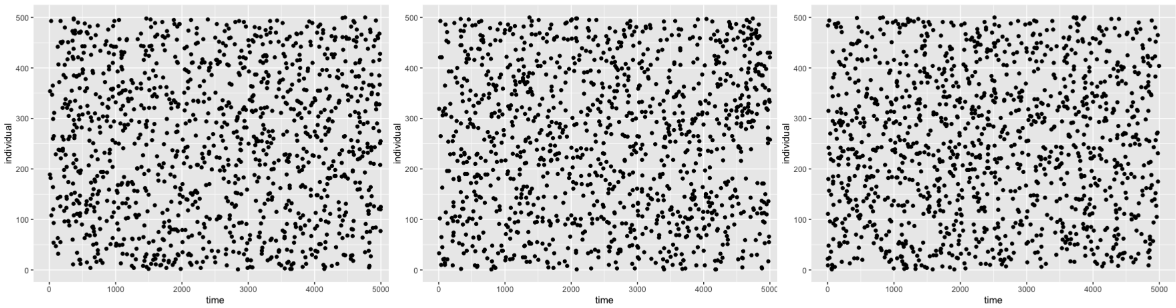

Figure 4: Adjacency plot of $Z_1 = X_1X_2$, $Z_2 = X_1(1-X_2)$, $Z_3 = (1-X_1)X_2$ when $X_{1,i,t}, X_{i,t}$ are independent, Bernoulli conditional on $q_{i,t}$, with $q_{i,t}$ i.i.d. across i and t following beta distribution.

5.2 Distribution of $n(S)$

As seen in Sections 3 and 4, the size $n(S)$ is a one-dimensional statistics that provides information on the underlying set distribution. The size is even a sufficient statistics in the exchangeable case. The pattern of the size distribution, in particular the number and location of its modes, can reveal the structure of positive/negative cross-sectional dependence between the underlying binary variables.

5.2.1 Case $n = 2$, $L = 1$: a toy example

To illustrate this effect, let us consider the case $n = 2$, $L = 1$. Then the distribution of (X_1, X_2) is characterized by the $(2, 2)$ contingency table: $\begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix}$. The set distribution is given by:

$$p(\emptyset) = p_{00}, \quad p(\{1\}) = p_{10} \quad p(\{2\}) = p_{01} \quad p(\{1,2\}) = p_{11},$$

and the distribution of the size is:

$$\mathbb{P}[n(S) = 0] = p_{00}, \quad \mathbb{P}[n(S) = 1] = p_{10} + p_{01}, \quad \mathbb{P}[n(S) = 2] = p_{11}.$$

We see that the distribution of the size identifies the probabilities of extreme events p_{00} and p_{11}, but partially identifies the probabilities p_{10} and p_{01} through their sum.

The exchangeable case arises when $\mathbb{P}[X_1 = 0] = \mathbb{P}[X_2 = 0]$, and $\mathbb{P}[X_1 = 1, X_2 = 0] = \mathbb{P}[X_1 = 0, X_2 = 1]$. These two conditions are equivalent to the single condition $p_{10} = p_{01}$. Then the contingency table becomes: $P = \begin{bmatrix} p_{00} & p_{01} \\ p_{01} & p_{11} \end{bmatrix}$, and the distribution of the size, given by:

$$\mathbb{P}[n(S) = 0] = p_{00}, \quad \mathbb{P}[n(S) = 1] = 2p_{01}, \quad \mathbb{P}[n(S) = 2] = p_{11},$$

identifies the underlying contingency table. The variables X_1, X_2 are positively (resp. negatively) dependent if and only if $\det P = p_{00}p_{11} - p_{01}^2 > 0$ (resp. < 0):

Proposition 21. *For $n = 2$ and exchangeable model:*

i) The positive dependence implies larger weights in extreme sizes, that is $p_{00} + p_{11} \geq 2p_{01}$, or equivalently $p_{01} \leq 1/4$.

ii) Large weight in middle size, i.e. $2p_{01} \geq p_{00} + p_{11}$, or equivalently $p_{01} \geq 1/4$, reveals negative dependence.

Proof. If the variables are positively dependent, we have:

$$2p_{01} - p_{00} - p_{11} \leq 2\sqrt{p_{00}p_{11}} - p_{00} - p_{11} = -\left(\sqrt{p_{00}} - \sqrt{p_{11}}\right)^2 \leq 0. \tag{4.3}$$

The result follows. □

Thus, for $n = 2$, a large mode at 1 reveals negative dependence, whereas positive dependence can create modes at extreme sizes.

The figures below provide the size distribution for 3 different sets of values of the $(2, 2)$ contingency table:

- In the first case, we take $\begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix} = \begin{bmatrix} 0.2 & 0.3 \\ 0.3 & 0.2 \end{bmatrix}$. We can check that this model has a LDPP representation, with $\Sigma = \begin{bmatrix} 1.5 & \sqrt{1.25} \\ \sqrt{1.25} & 1.5 \end{bmatrix}$, and the distribution of the size $n(S)$ is the convolution of two Bernoulli distributions with parameter 0.72 and 0.28, respectively.

- In the second case, we have[11] $\begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix} = \begin{bmatrix} 0.25 & 0.25 \\ 0.25 & 0.25 \end{bmatrix}$. It is easily checked that this model has a LDPP representation, with $\Sigma = \begin{bmatrix} 1 & 0.0 \\ 0 & 1 \end{bmatrix}$, so that all the principal minors $\det(\sigma_s)$ are equal to 1. By Proposition 13, the size $n(S)$ follows simply the binomial distribution $\mathcal{B}(2, 0.5)$.

- In the third case, we assume an exchangeable model where the stochastic intensity π follows the beta distribution with parameters $\alpha_1 = 0.36, \alpha_2 = 0.09$.

[11]This is a limiting case, where $p_{01} = 1/4$.

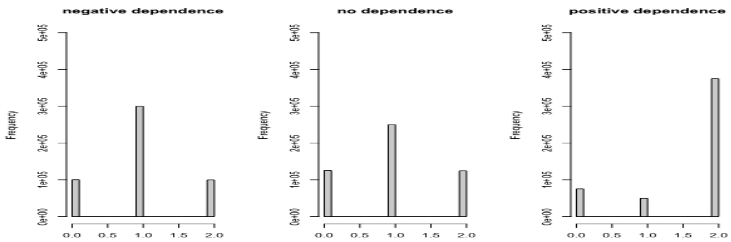

Figure 5: Three histograms of the size $n(S)$ for $n = 2$. Left panel is a LDPP model with negative dependence. Middle panel using a limiting LDPP model, with no dependence. The right panel is using an exchangeable model, leading to positive dependence. All histograms are computed based on a sampling size of $T = 50000$.

Let us provide a similar plot, but with a population of large size $n = 500$. In the four simulations below, we assume that $n(S)$ is the sum of:

i) a Bernoulli variable with stochastic intensity π, and a Binomial variable $\mathcal{B}(n-1, 0.025)$,

ii) a Binomial variable $\mathcal{B}(n/8, \pi)$ with stochastic intensity π, and a Binomial variable $\mathcal{B}(7n/8, 0.025)$,

iii) a Binomial variable $\mathcal{B}(n/4, \pi)$ with stochastic intensity π, and a Binomial variable $\mathcal{B}(3n/4, 0.025)$.

iv) a Binomial variable $\mathcal{B}(n/2, \pi)$ with stochastic intensity π, and a Binomial variable $\mathcal{B}(n/2, 0.025)$.

where the stochastic intensity π is equal to q with probability 0.1, and such that $\pi = 2q/3$ with probability 0.9, where q follows the beta distribution with parameters $\alpha_1 = 0.1$, $\alpha_2 = 0.01$.

The next figure provides the histogram of the size $n(S)$ under these four data generating processes.

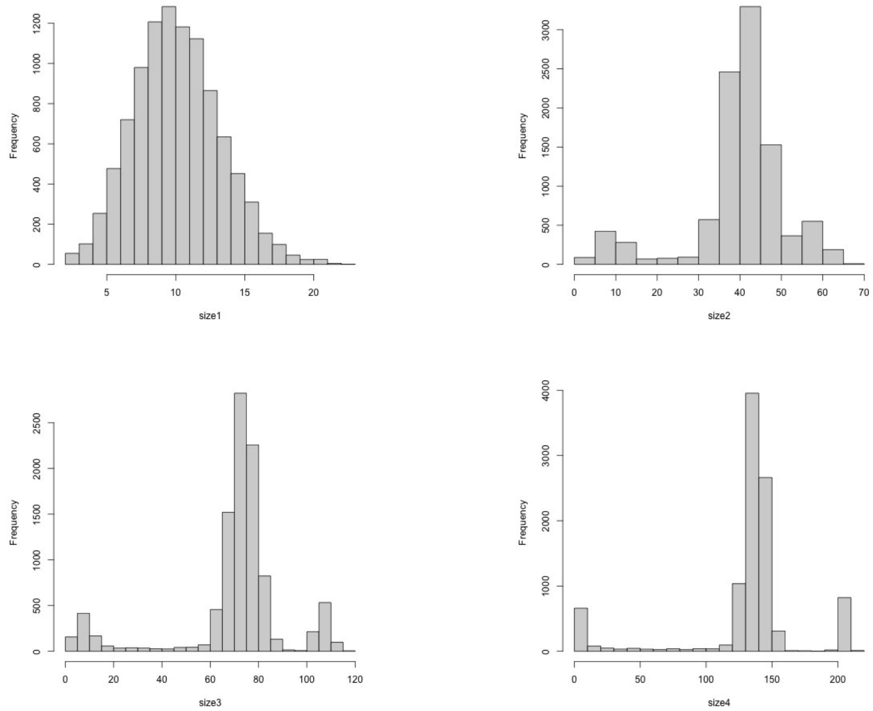

Figure 6: Histograms of the four distributions of $n(S)$ with a sample size of T =20000. Northwest panel: distribution i). Northeast panel: distribution ii). Southwest panel: distribution iii). Southeast panel: distribution iv).

In case i), when the size of the binomial variable with stochastic intensity is only 1, the distribution (northwest panel) is unimodal. This is the effect of repulsion. When the size of the binomial variable with stochastic intensity increases, three modes appear, and around these three modes, the masses become more and more concentrated.

5.3 Distribution of the size of hitting sets

The distribution of size $n(S)$ depends on n independent parameters, that correspond to the elementary probabilities, whereas the distribution of the set depends on $2^n - 1$ independent parameters. Therefore the distribution of the size cannot be very informative in an unconstrained framework, with a degree of under-identification equal to $2^n - 1 - n$.

It can be interesting to complete the analysis by considering other multivariate definitions of the size. Let us consider a partition $s_1, \cdots, s_L$ of $\{1, ..., n\}$:

$$\{1, ..., n\} = s_1 \cup \cdots \cup s_L, \qquad \text{with } s_j \cap s_k = \emptyset, \qquad \forall j \neq k.$$

Then we have:

$$n(S) = n(S \cap s_1) + \cdots + n(S \cap s_L).$$

Instead of the distribution of $n(S)$, we can now consider the joint distribution of $\Big[n(S \cap s_1), \cdots, n(S \cap s_L)\Big]$, for different partitions.

As an illustration, We take $n = 400$, $L = 2$, and use the same kernel as in Figure 4: $K^+ = 0.02Id + 0.0002J$.

$$s_1 = \{1, 2, ..., 200\}, \quad s_2 = \{201, 102, ..., 400\}$$

Because this LDPP model is exchangeable and s_1, s_2 have the same size, the sizes of the two hitting sets $n(S \cap s_k), k = 1, 2$ are also exchangeable. Both follow the size distribution of a LDPP, with kernel $K^+_{s_1}$ and $K^+_{s_2}$, respectively. Since K^+ is exchangeable, we have $K^+_{s_1} = K^+_{s_2}$. We use the LDPP sampling algorithm proposed by Kulesza and Taskar (2012) and Launay et al. (2020) to simulate $M = 50000$ realizations of the LDPP, and compute the resulting size of the hitting sets $n(S \cap s_k), k = 1, 2$. Their empirical correlation coefficient is -0.0352, which echoes the repulsive property of the LDPP. We plot below the empirical histograms and the joint density of the sizes of the two hitting sets.

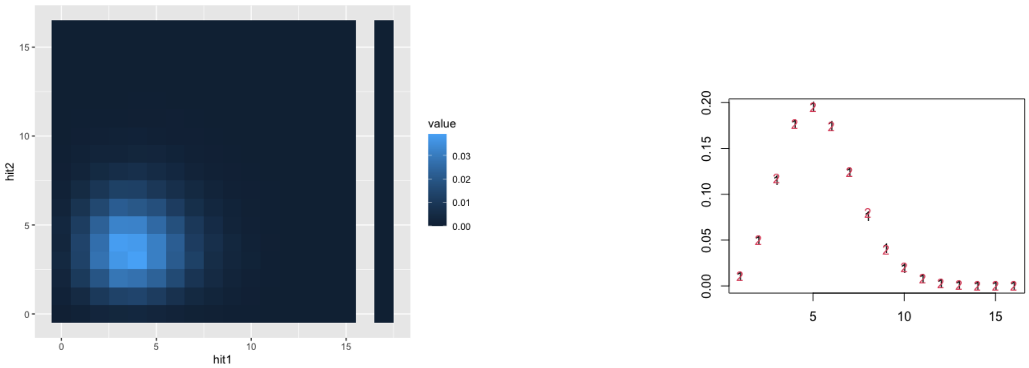

Figure 7: Left panel: heat map of the joint distribution of the sizes of the two hitting sets in an exchangeable LDPP model. Right panel: marginal histogram of $n(S \cap s_1)$ and $n(S \cap s_2)$ Both histograms are very close since they converge to the same distribution when M goes to infinity.

Next, we consider a non exchangeable LDPP. We simulate a matrix Σ from the Wishart distribution with degree of freedom 3 and matrix parameter diag$(1, ..., n)$. We rank the diagonal terms of Σ in decreasing order, and denote by s_1 (resp. s_2) the set of locations i such that Σ_{ii} is among the first two quartiles among all the diagonal terms. Then we simulate M independent copies of S_t from the LDPP model with matrix Σ. Figure 9 below is the analogue of Figure 8. It shows that the distributions of $n(S_t \cap s_1)$ and of $n(S_t \cap s_2)$ are significantly different.

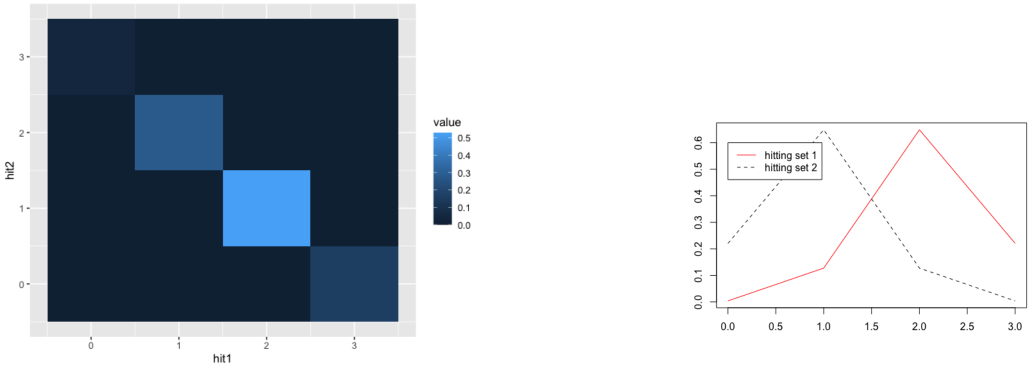

Figure 8: Left panel: heat map of the joint distribution of the two hitting sets in a non exchangeable LDPP model. Right panel: marginal distribution function of $n(S \cap s_1)$ and $n(S \cap s_2)$. The two marginal distributions are significantly different.

5.4 Alternative variance-covariance matrix

The aim of this section is to discuss different notions of independence and their interpretation. Let us consider a partition of $\{1, ..., n\} = s_1 + \cdots + s_L$ into L subsets and denote $Z_l = \prod_{i \in s_l} X_i$, $l = 1, ..., L$. Thus Z_l is a binary variable with value 1, if and only if all the X_i's in s_l take value 1. We have the following property:

Proposition 22. *Let us consider the $(L \times L)$ matrix Γ with elements:*

$$\gamma_{j,l} = G^+(s_j \cup s_l) - G^+(s_j)G^+(s_l), \qquad j, l = 1, ..., L.$$

Then the matrix Γ is the variance-covariance matrix of the vector Z of binary variables.

Proof. This is a consequence of $Z_l = 1$ iff $X_i = 1, \forall i \in s_l$. □

For instance, if $s_1 = \{1\}$ and $s_2 = \{2, 3\}$, we have:

$$\begin{aligned}\gamma_{1,2} &= \mathbb{P}[1, 2, 3 \in S] - \mathbb{P}[1 \in S]\mathbb{P}[2, 3 \in S] \\ &= \mathbb{E}[X_1 X_2 X_3] - \mathbb{E}[X_1]\mathbb{E}[X_2 X_3] \\ &= \text{Cov}[X_1, X_2 X_3] = \text{Cov}[Z_1, Z_2].\end{aligned}$$

When L is large, this matrix can have a large dimension, but can be represented by a coloured plot to disentangle its negative, zero, and positive values[12].

As an illustration, we start with an LDPP model with factor representation, in which factor β is such that β_i is proportional to i, $i = 1, ..., n$. We choose $n = 400$, $\sigma = 0.01$, $\lambda = 0.5(1^2 + 2^2 +$

[12]or by an adjacency plot with cross for positive values, points for negative values.

$\cdots + n^2$). Most of the eigenvalues of its kernel K^+ are close to zero, except the largest, which is very close to 1. Thus, with a large probability, the size of this LDPP is equal to 1. This would lead to a situation where variables Z_l are nearly always equal to zero. Thus we alternatively consider the complement of this LDPP, which, by Rising (2013) Th.2.3.5, is also an LDPP with kernel $Id - K^+$. Then we consider $L = 10$ subsets, with

$$s_1 = \{1, ..., n/10\}, s_2 = \{1 + n/10, 2 + n/10, \cdots, 2n/10\}, \cdots s_{10} = \{9n/10 + 1, \cdots, n\}.$$

We simulate a total of M realizations from this complementary LDPP, compute $Z_1, Z_2, ..., Z_{10}$. We represent its correlation matrix in the left panel of Figure 10 below. As a comparison, we also plot the $(n \times n)$ correlation matrix between the X_i's, $i = 1, .., n$, and compare it to the correlation matrix from an exchangeable Bernoulli model with stochastic intensity.

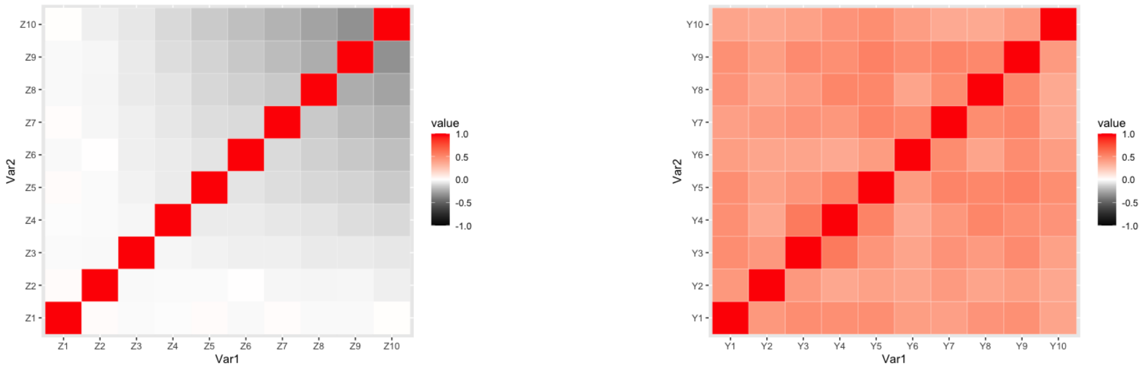

Figure 9: Left panel: correlation matrix of the LDPP model whose matrix Σ follows Wishart distribution with degree of freedom $n+1$ and matrix parameter $0.01Id$. Right panel: correlation matrix of an exchangeable LDPP model where X_i follows $\mathcal{B}(1, \pi)$ and π follows the beta distribution with parameters $\alpha_1 = \alpha_2 = 1$.

In the left panel, correlations are mostly negative, due to the repulsive property of the LDPP. Moreover, the negative correlation is especially pronounced between Z_8, Z_9, Z_{10}. This is due to the fact that the factor β is chosen such that β_i is dominant for large index i. On the right panel, correlations are mostly positive, with comparable values, due to the exchangeability property.

To understand how to use such information, let us consider partitions into two sets s_1 and $s_2 = \bar{s_1}$. Then we can compute the associated correlation: $\rho(s_1) = \text{cor}(Z_1, Z_2) := \gamma_{12}/\sqrt{\gamma_{11}\gamma_{22}}$.

This correlation is equal to 1, if and only if $Z_2 = Z_1$, equal to -1, if and only if $Z_2 = 1 - Z_1$. In the first case, $Z_1 = 1$ means that we have high risk for all individuals in s (i.e. a type of cluster of risks) and then at the same time high risk for all individuals in $\bar{s_1}$. This is the reverse situation for correlation equal to -1.

Then, by analogy with principal component analysis, we can look for

$$ms^* = \arg\min_s \rho(s), \qquad \text{and} \qquad Ms^* = \arg\max_s \rho(s).$$

5.5 Updating the tranche prices

The prices of basket derivatives have been computed without additional information on the random set. This price has to be updated if additional information is available during the year of observation. As an example, let us consider a tranche (s_1, s_2), with $s_1 \subset s_2$, and assume that in the middle of the year we already observe a set s_0 of the revealed risks, then the computation has to be updated and performed conditional on $S \supset s_0$.

Let us for instance consider the semi-up tranche with $s_1 = \{2\}$ and $s_2 = \{1, ..., n\}$, that is, this derivative pays 1 \$, if and only if $2 \in S$. By Section 3.3, its price is $Q^+(\{2\}) = \mathbb{Q}[2 \in S]$. Let us now assume that we have the additional knowledge that $1 \in S$, then the updated new price is $\mathbb{Q}[2 \in S | 1 \in S]$. This price will in general be different from the initial price. Let us illustrate this updating using two examples.

Example 6. We assume that S follows a LDPP under the risk-neutral probability $\mathbb{Q}$. Then we have: $\mathbb{Q}[2 \in S] = (K^+)_{2,2}$, where $K^+ = \Sigma(Id + \Sigma)^{-1}$. Moreover, by Proposition 18, the distribution of $S - \{1\} | 1 \in S$ is still a $(n-1)$-dimensional LDPP, thus $\mathbb{Q}[2 \in S | 1 \in S] = \Big[K^+(\{1\})\Big]_{2,2}$, where $K^+(\{1\})$ is defined by eq. (3.14).

Numerically, if $n = 3$, and $\Sigma = 0.5Id + J$, we get:

$$\mathbb{Q}[2 \in S] = 0.48, \qquad \mathbb{Q}[2 \in S | 1 \in S] = 0.43.$$

In other words, the updated tranche price is lower than its initial price, due to the repulsive property of the LDPP.

Example 7. We now assume that $n = 3$, and X_1, X_2, X_3 are conditionally i.i.d. Bernoulli given a stochastic probability parameter q, where q follows the beta distribution with parameters $\alpha_1 = \alpha_2 = 2$. Then by the beta-Bernoulli conjugacy, we have:

$$\mathbb{Q}[2 \in S] = \frac{\alpha_1}{\alpha_1 + \alpha_2} = 0.5, \qquad \mathbb{Q}[2 \in S | 1 \in S] = \frac{\alpha_1 + 1}{\alpha_1 + \alpha_2 + 1} = 0.6.$$

In other words, because of the positive dependence among X_i's, the updated price is higher than the initial price.

What is the effect of conditioning on the stochastic dominance? That is, if S right dominates S^* at order 1, and if s_0 is a subset of, say, $\{1, ..., k\}$, where $k < n$, then would the conditional distribution of $S \cap \{k+1,...,n\}$ given $S \cap \{1,...,k\} = s_0$ still right dominates the conditional distribution of $S^* \cap \{k+1,...,n\}$ given $S^* \cap \{1,...,k\} = s_0$? In other words, will the semi-up tranche price still be higher for S than for S after conditioning?

We can check that these two conditional distributions have conditional pmf's and conditional left cumulated functions:

$$\frac{p(s_0 \cup s)}{G^-(s_0)}, \quad \frac{p^*(s_0 \cup s)}{G^{*-}(s_0)}, \text{ for pmf's}$$
$$\frac{G^-(s_0 \cup s)}{G^+(s_0)}, \quad \frac{G^{*-}(s_0 \cup s)}{G^{*-}(s_0)}, \text{ for left cumulated functions.}$$

respectively, for any subset $s \subset \cap\{k+1,...,n\}$. Because S right dominates S^* at order 1, both the denominator and the numerator of $\frac{G^-(s_0 \cup s)}{G^+(s_0)}$ dominate their counterpart for S^*. Thus, the stochastic dominance is not necessarily preserved after the conditioning.

6 Concluding remarks

The objective of this paper was to define the notions of stochastic dominances at order 1 and 2 for random sets in a finite space, that is when several binary variables are jointly observed. In the random set framework, the left and right definitions of stochastic dominance differ, even in special cases as for the LDPP family. This analysis leads to different measures of risk on random sets and to the introduction of well-designed basket (i.e. set) derivatives to hedge against these risks. We especially discuss tranches and European calls written on random sets and their closed form prediction and pricing formulas.

The pricing formulas have been derived for panel data under the assumption of i.i.d. observations of random sets. Thus we allow for cross-sectional dependence, but not for serial dependence. The main results of the paper can be extended to also include serial dependence [see e.g. Gouriéroux and Lu (2023b) for Markov LDPP models and Gouriéroux and Lu (2023a) for dynamic LDPP models with Wishart autoregressive stochastic kernel].

References

Affandi, R., Kulesza, A., and Fox, E. (2012). Markov Determinantal Point Processes. In *Proceedings of the Twenty-Eighth Conference on Uncertainty in Artificial Intelligence*, pages 26–35.

Aumann, R. J. (1965). Integrals of Set-Valued Functions. *Journal of Mathematical Analysis and Applications*, 12(1):1–12.

Beresteanu, A., Molchanov, I., and Molinari, F. (2011). Sharp Identification Regions in Models with Convex Moment Predictions. *Econometrica*, 79(6):1785–1821.

Borodin, A. and Rains, E. M. (2005). Eynard–Mehta Theorem, Schur Process, and their Pfaffian Analogs. *Journal of Statistical Physics*, 121:291–317.

Bromwich, T. (2005). *An Introduction to the Theory of Infinite Series*, volume 335. American Mathematical Society.

Brunel, V.-E. (2018). Learning Signed Determinantal Point Processes Through the Principal Minor Assignment Problem. *Advances in Neural Information Processing Systems*, 31.

Brunel, V.-E., Moitra, A., Rigollet, P., and Urschel, J. (2017). Rates of Estimation for Determinantal Point Processes. *Proceedings of Machine Learning Research*, 65:1–3.

Bryzgalova, S., Lerner, S., Lettau, M., and Pelger, M. (2022). Missing Financial Data. *Available at SSRN 4106794*.

Debreu, G. (1967). Integration of Correspondences. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Le Cam, L., and J., Neyman, eds*, volume 2, pages 351–372. University of California Press Berkeley, CA.

Denuit, M., Eeckhoudt, L., Tsetlin, I., and Winkler, R. (2013). Multivariate Concave and Convex Stochastic Dominance. *In: Biagini, F., Richter, A., Schlesinger, H. (eds) Risk Measures and Attitudes. EAA Series, Springer London, p11-32.*

Denuit, M. M. and Mesfioui, M. (2010). Generalized Increasing Convex and Directionally Convex Orders. *Journal of Applied Probability*, 47(1):264–276.

Duffie, D., Pan, J., and Singleton, K. (2000). Transform Analysis and Asset Pricing for Affine Jump-Diffusions. *Econometrica*, 68(6):1343–1376.

Fishburn, P. (1974). Convex Stochastic Dominance with Continuous Distribution Functions. *Journal of Economic Theory*, 7(2):143–158.

Fishburn, P. and Vickson, R. (1978). Theoretical Foundations of Stochastic Dominance. In *Stochastic Dominance: An Approach to Decision-Making Under Risk, Whitmore, GA and Findlay, M (eds.)* . DC Health London.

Gouriéroux, C. and Lu, Y. (2023a). Determinantal Point Process with Stochastic Intensity. *CREST DP*.

Gouriéroux, C. and Lu, Y. (2023b). Markov Determinantal Point Processes. *CREST DP*.

Griffin, K. and Tsatsomeros, M. J. (2006). Principal Minors, Part II: The Principal Minor Assignment Problem. *Linear Algebra and its Applications*, 419(1):125–171.

Hanoch, G. and Levy, H. (1969). The Efficiency Analysis of Choices Involving Risk. *Review of Economic Studies*, 36(3):335–346.

Heath, D. and Sudderth, W. (1976). De Finetti's Theorem on Exchangeable Variables. *The American Statistician*, 30(4):188–189.

Henderson, H. and Searle, S. (1979). Vec and Vech Operators for Matrices, with Some Uses in Jacobians and Multivariate Statistics. *Canadian Journal of Statistics*, 7(1):65–81.

Hough, J., Krishnapur, M., Peres, Y., and Virág, B. (2006). Determinantal Processes and Independence. *Probability Surveys*, 3:206–229.

Kamae, T., Krengel, U., and O'Brien, G. (1977). Stochastic Inequalities on Partially Ordered Spaces. *Annals of Probability*, 5(6):899–912.

Kihlstrom, R. and Mirman, L. (1974). Risk Aversion with Many Commodities. *Journal of Economic Theory*, 8(3):361–388.

Kulesza, A. and Taskar, B. (2012). Determinantal Point Processes for Machine Learning. *Foundations and Trends in Machine Learning*, 5(2–3):123–286.

Launay, C., Galerne, B., and Desolneux, A. (2020). Exact Sampling of Determinantal Point Processes Without Eigendecomposition. *Journal of Applied Probability*, 57(4):1198–1221.

Leshno, M., Levy, H., and Spector, Y. (1997). A Comment on Rothschild and Stiglitz's "Increasing Risk: I. A Definition". *Journal of Economic Theory*, 77(1):223–228.

Marshall, A. W. (1991). Multivariate Stochastic Orderings and Generating Cones of Functions. In *Stochastic Orders and Decision under Risk, IMS Lecture Notes-Monograph Series, K. Mosler, M. Scarsini (eds).*, volume 19, pages 231–247. Institute of Mathematical Statistics Lecture Notes.

Matheron, G. (1975). *Random Sets and Integral Geometry.* Wiley, New-York.

Molchanov, I. (2017). *Theory of Random Sets, 2nd Edition.* Springer.

Molchanov, I. and Molinari, F. (2018). *Random Sets in Econometrics.* Econometric Society Monographs, Cambridge University Press.

Molchanov, I. and Schmutz, M. (2010). Multivariate Extension of Put-Call Symmetry. *SIAM Journal on Financial Mathematics*, 1(1):396–426.

Molchanov, I. and Schmutz, M. (2014). Multiasset Derivatives and Joint Distributions of Asset Prices. *In: Kabanov, Y., Rutkowski, M., Zariphopoulou, T. (eds) Inspired by Finance. Springer, Cham. p439–459.*

Mosler, K. and Scarsini, M. (1991). Some Theory of Stochastic Dominance. In *Stochastic Orders and Decision under Risk, IMS Lecture Notes-Monograph Series, K. Mosler, M. Scarsini (eds), volume 19, 261-284*. Institute of Mathematical Statistics Lecture Notes.

Müller, A. (1997). Stochastic Orders Generated by Integrals: A Unified Study. *Advances in Applied Probability*, 29(2):414–428.

Perez, L. (2018). Modularity Classes and Boundary Effects in Multivariate Stochastic Dominance. *arXiv preprint arXiv:1807.06402*.

Rising, J. K. (2013). *Advances in the Theory of Determinantal Point Processes*. PhD thesis, University of Pennsylvania.

Rothschild, M. and Stiglitz, J. (1970). Increasing Risk: I. A Definition. *Journal of Economic Theory*, 2(3):225–243.

Scarsini, M. (1988). Dominance Conditions for Multivariate Utility Functions. *Management Science*, 34(4):454–460.

Shaked, M. and Shanthikumar, G. (2007). *Stochastic Orders*. Springer.

Strassen, V. (1965). The Existence of Probability Measures with Given Marginals. *Annals of Mathematical Statistics*, 36(2):423–439.

Tsetlin, I. and Winkler, R. L. (2018). Multivariate Almost Stochastic Dominance. *Journal of Risk and Insurance*, 85(2):431–445.

Vickson, R. (1975). Stochastic Dominance for Decreasing Absolute Risk Aversion. *Journal of Financial and Quantitative Analysis*, 10(5):799–811.

Appendix A Proofs of Propositions

Appendix A.1 Proof of Proposition 1 ($L = 1$)

As noted in the text, Proposition 1 can be derived by the Moebius inversion formula (2.3). We provide below an algorithmic proof of uniqueness, that can be more appropriate numerically to derive function v from function V^+.

Let us consider a $S-$decreasing function V^+ and explain how to derive v in a unique way. We start from the maximal element $s = \{1, ..., n\}$. We have:

$$V^+(\{1, 2, ..., n\}) = v(\{1, 2, ..., n\}).$$

Thus $v(\{1, 2, ..., n\})$ is known. Let us now consider the complement of $\{i\}$ in $\{1, 2, ..., n\}$, i.e. $\{1, 2, ..., n\} - \{i\}$. We have:

$$v(\{1, 2, ..., n\} - \{i\}) = V^+(\{1, 2, ..., n\} - \{i\}) - V^+(\{1, 2, ..., n\}).$$

Hence $v(\{1, 2, ..., n\} - \{i\})$ is known for any i. Next we find the complement of $\{i, j\}$ in $\{1, 2, ..., n\}$, and continue to go down on the ordering tree.

If $L > 1$, we can use the bijection between $\Big[\mathcal{P}\{1, ..., n\}\Big]^L$ and $\mathcal{P}\Big(\{1, ..., n\} \times \{1, ..., l\}\Big)$ to transform the problem to the case where $L = 1$.

Appendix A.2 Proof of Proposition 3

We have:

$$\begin{aligned}
\sum_s v(s)W^+(s) &= \sum_s \sum_{s^* \supset s} v(s)w(s^*) \\
&= \sum_s \sum_{s^* \subset s} w(s)v(s^*) \\
&= \sum_s \big[w(s) \sum_{s \supset s^*} v(s^*)\big] \\
&= \sum_s w(s)V^-(s).
\end{aligned}$$

Appendix B The Left Cumulated Functions

Appendix B.1 The LDPP family

In Proposition 4, we have computed the expression of function G^+ in the LDPP family. Let us now compute G^- in the same model. We have:

Proposition 23.

$$G^-(s) = \det K^-_{\bar{s}}, \qquad \text{(eq. a.4)}$$

where $\bar{s}$ is the complement of s and matrix K^- is defined by:

$$K^- = Id - K^+ = (Id + \Sigma)^{-1}. \qquad \text{(eq. a.5)}$$

Proof. By Rising (2013), Theorem 2.3.5, when S follows a LDPP, its complement follows also a LDPP with kernel $K^- = Id - K^+$. Then we have:

$$
\begin{aligned}
G^-(s) &= \sum_{s^* \subset s} p(s^*) = \sum_{s^* \subset s} \mathbb{P}(S = s^*) \\
&= \sum_{\bar{s} \subset \bar{s^*}} \mathbb{P}(\bar{S} = \bar{s^*}) \\
&= \det(K^-_{\bar{s}}).
\end{aligned}
$$

□

More generally, we have:

$$
G^{(p)-}(s) = \det \left[(p-1) I_s + K^- \right].
$$

The matrices K^- and Σ are also in a one-to-one relationship, since we have:

$$
\Sigma = (K^-)^{-1} - Id. \qquad \text{(eq. a.6)}
$$

Appendix B.2 The left stochastic dominance

Appendix B.2.1 Definition

In the standard case of stochastic dominance based on a total order, it is equivalent to define the dominance at order 1 from the survival function or from the c.d.f.. This is no longer the case with partial order. By analogy with Definition 3, we could have defined the stochastic dominance at order 1 as follows:

Definition 9 (Left dominance at order 1)**.** S^* left dominates S at order 1 if and only if $G^{*-}(s) \leq G^-(s), \quad \forall s$.

Appendix B.2.2 Non equivalence between left and right dominances in the general case

It is easily checked that the left and right dominances at order 1 do not define the same partial order. Let us consider the case $n = 2$ and write the conditions for stochastic dominance.

- right stochastic dominance:

$$
\begin{aligned}
p^*(\{1,2\}) &\geq p(\{1,2\}), && \text{(eq. a.7)} \\
p^*(\{1\}) + p^*(\{1,2\}) &\geq p(\{1\}) + p(\{1,2\}), && \text{(eq. a.8)} \\
p^*(\{2\}) + p^*(\{1,2\}) &\geq p(\{2\}) + p(\{1,2\}). && \text{(eq. a.9)}
\end{aligned}
$$

- left stochastic dominance:

$$\begin{aligned} p^*(\emptyset) &\leq p(\emptyset), \\ p^*(\emptyset) + p^*(\{1\}) &\leq p(\emptyset) + p(\{1\}), \\ p^*(\emptyset) + p^*(\{2\}) &\leq p(\emptyset) + p(\{2\}). \end{aligned}$$

These inequalities are equivalent to:

$$p^*(\{1\}) + p^*(\{2\}) + p^*(\{1,2\}) \geq p(\{1\}) + p(\{2\}) + p(\{1,2\}), \quad \text{(eq. a.10)}$$

$$p^*(\{1\}) + p^*(\{1,2\}) \geq p(\{1\}) + p(\{1,2\}), \quad \text{(eq. a.11)}$$

$$p^*(\{2\}) + p^*(\{1,2\}) \geq p(\{2\}) + p(\{1,2\}), \quad \text{(eq. a.12)}$$

since the elementary probabilities sum up to 1. We see that equations (eq. a.7)-(eq. a.9) do not imply equations (eq. a.10)-(eq. a.12), and vice versa. In other words, none the right and left stochastic dominance implies the other one when $n = 2$.

For instance, if S^* and S are such that:

$$p^*(\{1\}) = 0.5, \quad p^*(\{2\}) = 0.4, \quad p^*(\{1,2\}) = 0.05, \quad p^*(\emptyset) = 0.05 \quad \text{(eq. a.13)}$$

$$p(\{1\}) = 0.4, \quad p(\{2\}) = 0.3, \quad p(\{1,2\}) = 0.1, \quad p(\emptyset) = 0.2 \quad \text{(eq. a.14)}$$

then S^* left dominates S, but does not right dominate S, since (eq. a.7) is not satisfied.

Similarly, if S^* and S are such that:

$$p^*(\{1\}) = 0.2, \quad p^*(\{2\}) = 0.3, \quad p^*(\{1,2\}) = 0.3, \quad p^*(\emptyset) = 0.2 \quad \text{(eq. a.15)}$$

$$p(\{1\}) = 0.4, \quad p(\{2\}) = 0.4, \quad p(\{1,2\}) = 0.1, \quad p(\emptyset) = 0.1 \quad \text{(eq. a.16)}$$

then S^* right dominates S, but does not left dominate S, since (eq. a.10) is not satisfied.

Appendix B.2.3 Non equivalence between left and right dominances at order 1 in the LDPP framework

Is it possible for the left and right dominances at order 1 to coincide, under the extra LDPP assumption? Let us recall that by Proposition 11, the right dominance at order 1 of S^* over S is equivalent to:

$$\det(K_s^{*+}) \geq \det(K_s^{+}), \qquad \forall s \subset \{1, ..., n\}. \quad \text{(eq. a.17)}$$

Similarly, by Proposition 26, the left dominance at order 1 of S^* over S is equivalent to:

$$\det(K_s^{*-}) \geq \det(K_s^-), \qquad \forall s \subset \{1, ..., n\}. \qquad \text{(eq. a.18)}$$

where $K^{*-} = Id - K^{*+}$ and $K^- = Id - K^+$. Because generally $\det(Id_s - K_s^{*+}) + \det(K_s^{*+})$ is not equal to 1, neither of (eq. a.17) and (eq. a.18) implies the other.

As a counterexample, we can remark that the distributions of S^* and S defined in (eq. a.13), (eq. a.14) can be written as LDPP, with:

$$K^{*+} = \begin{bmatrix} 0.5 + 0.05 & \sqrt{(0.5 + 0.05)(0.4 + 0.05) - 0.05^2} \\ * & 0.4 + 0.05 \end{bmatrix}, \qquad K^+ = \begin{bmatrix} 0.4 + 0.1 & \sqrt{(0.4 + 0.1)(0.3 + 0.1) - 0.1^2} \\ * & 0.3 + 0.1 \end{bmatrix}.$$

Similarly, the distributions of S^* and S defined in (eq. a.15), (eq. a.16) can also be written as LDPP, with:

$$K^{*+} = \begin{bmatrix} 0.2 + 0.3 & \sqrt{(0.2 + 0.3)(0.2 + 0.3) - 0.3^2} \\ * & 0.3 + 0.3 \end{bmatrix}, \qquad K^+ = \begin{bmatrix} 0.4 + 0.1 & \sqrt{(0.4 + 0.1)(0.4 + 0.1) - 0.1^2} \\ * & 0.4 + 0.1 \end{bmatrix}.$$

Appendix C LDPP Family with Factor Representation

Appendix C.1 Definition

The model assumes that $\Sigma = \sigma^2 Id + \lambda\beta\beta'$, where $\sigma^2 \geq 0, \lambda \geq 0$ and β is a vector of unit norm $\beta'\beta = 1$. Thus, the underlying dependence has rank 1 and is driven by β. We have:

$$\big(Id + \Sigma\big)^{-1} = \frac{1}{1 + \sigma^2} Id - \frac{\lambda}{(1 + \sigma^2)(1 + \lambda + \sigma^2)} \beta\beta',$$

and $K^+ = \frac{\sigma^2}{1+\sigma^2} Id + \frac{\lambda}{(1+\sigma^2)(1+\lambda+\sigma^2)} \beta\beta'$, by (eq. a.5). Thus the marginal kernel has also a factor representation with the same factor β.

This factor representation includes the special case of equicorrelation in Σ (or equivalently K^+), corresponding to $\beta = \frac{1}{\sqrt{n}} \mathbb{1}$, where $\mathbb{1}$ is the vector with components equal to 1.

Appendix C.2 Subkernels

Then the subkernels K_s^+ have also a factor representation with subfactor directions. We have:

$$K_s^+ = \frac{\sigma^2}{1 + \sigma^2} Id_s + \frac{\lambda \beta_s' \beta_s}{(1 + \sigma^2)(1 + \lambda + \sigma^2)} \beta(s)\beta(s)',$$

where $\beta(s) = \beta_s / \sqrt{\beta_s' \beta_s}$. For subkernels, there is one factor if $\beta(s) \neq 0$, no factor, otherwise.

Appendix C.3 Eigenvalues

We deduce the eigenvalues of the different matrices of interest:

- for Σ, σ^2 with multiplicity order $n-1$, $\sigma^2+\lambda$, with order 1.
- for K^+, $\frac{\sigma^2}{1+\sigma^2}$ with multiplicity order $n-1$, $\frac{\lambda+\sigma^2}{1+\lambda+\sigma^2}$, with order 1.
- for K_s^+, $\frac{\sigma^2}{1+\sigma^2}$ with multiplicity order $n(s)-1$, $\frac{\sigma^2}{1+\sigma^2}+\frac{\lambda\beta_s'\beta_s}{(1+\sigma^2)(1+\lambda+\sigma^2)}$, with order 1.

Note that $\frac{\sigma^2}{1+\sigma^2}+\frac{\lambda}{(1+\sigma^2)(1+\lambda+\sigma^2)}=\frac{\lambda+\sigma^2}{1+\lambda+\sigma^2}$.

Appendix C.4 Distribution

From the eigenvalues of K_s^+, we deduce the survival function of random set S:

$$G^+(s)=\frac{\sigma^{2[n(s)-1]}}{(1+\sigma^2)^{n(s)}}[\sigma^2+\frac{\lambda\beta_s'\beta_s}{1+\lambda+\sigma^2}],\text{ if } s\neq\emptyset,\quad G^+(\emptyset)=1.$$

For instance, if $s=\{i\}$ is a singleton, we get: $G^+(\{i\})=\mathbb{P}[X_i=1]=\frac{\sigma^2}{1+\sigma^2}+\frac{\lambda\beta_i^2}{(1+\sigma^2)(1+\lambda+\sigma^2)}$, which is increasing in $|\beta_i|$. In particular, in the special case where $\sigma=0$, Σ is of rank 1. Hence S is of size 0 or 1 almost surely and $\mathbb{P}[X_i=1]=\frac{\lambda\beta_i^2}{1+\lambda}$. Because the components of the vector $(\beta_1^2,...,\beta_n^2)$ sum up to unity, it is the vector of probabilities of the multinomial distribution of $(X_1,...,X_n)$, given that $X_1+\cdots+X_n=1$, and $\frac{\lambda}{1+\lambda}$ (resp. $\frac{0}{1+\lambda}$) is the probability that $X_1+\cdots+X_n=1$ (resp. $X_1+\cdots+X_n=0$).

From the eigenvalues of K^+, we deduce that the distribution of $n(S)$ is the convolution of the binomial distribution $\mathcal{B}(n-1,\frac{\sigma^2}{1+\sigma^2})$ and the Bernoulli distribution $\mathcal{B}(1,\frac{\lambda+\sigma^2}{1+\lambda+\sigma^2})$. Thus we have:

$$\mathbb{E}[n(S)]=(n-1)\frac{\sigma^2}{1+\sigma^2}+\frac{\lambda+\sigma^2}{1+\lambda+\sigma^2},$$
$$\mathbb{V}[n(S)]=(n-1)\frac{\sigma^2}{(1+\sigma^2)^2}+\frac{\lambda+\sigma^2}{(1+\lambda+\sigma^2)^2}.$$

In particular, the parameters λ,σ^2 are chosen with a given expected mean m, if they satisfy:

$$(n-1)\frac{\sigma^2}{1+\sigma^2}+\frac{\sigma^2+\lambda}{1+\sigma^2+\lambda}=m.$$

Online Appendix 1: Links between the distributions of X and S

Let us denote $\bar{X}_i=1-X_i, i=1,...,n$, and $\bar{S}$ the complement of S in $\{1,...,n\}$. We have the following definitions and relationships.

1. $p(s) = \mathbb{P}[S = s] = \mathbb{P}[X_i = 1, \text{ if } i \in s, X_i = 0, \text{ if } i \in \bar{s}]$.

2. $G^+(s) = \mathbb{P}[S \supset s] = \mathbb{P}[X_i = 1, \text{ if } i \in s]$. In particular $G^+(\{i\}) = \mathbb{P}[X_i = 1]$ defines the marginal distribution of X_i, $G^+(\{i,j\}) = \mathbb{P}[X_i = 1, X_j = 1]$ defines the pairwise distribution of X_i, X_j, once their marginal distributions are known.

3. $G^+(\{i,j\}) - G^+(\{i\})G^+(\{j\}) = \text{Cov}(X_i, X_j)$.

4. $\bar{p}(\bar{s}) := \mathbb{P}[\bar{S} = \bar{s}] = \mathbb{P}[S = s] = \mathbb{P}[\bar{X}_i = 1, \text{ if } i \in \bar{s}, X_i = 0, \text{ if } i \in s]$.

5. $\bar{G}^+(\bar{s}) := \mathbb{P}[\bar{S} \supset \bar{s}] = \mathbb{P}[S \subset s] = G^-(s)$.

Online Appendix 2: Laplace transform

OA. 2.1. Definition

Let us assume $L = 1$. The Laplace transform of the set of binary variables $(X_1, ..., X_n)$ is defined as:

$$\Psi_X(u) = \mathbb{E}\Big[\exp\big(-u_1X_1 - u_2X_2 - \cdots - u_nX_n\big)\Big], \qquad \text{(eq. a.19)}$$

where $u_i \geq 0, i = 1, ..., n$. This Laplace transform can also be written in terms of the random set S. Let us denote:

$$u(s) = \sum_{i \in s} u_i, \qquad \text{(eq. a.20)}$$

with the convention $\sum_{i \in \emptyset} u_i = 0$. The function $u(\cdot)$ is a positive increasing function of the set-valued argument s. Then equation (eq. a.19) becomes:

$$\Psi_S(u) = \mathbb{E}[\exp(-u(S))]\Big(= \Psi_X(u)\Big). \qquad \text{(eq. a.21)}$$

OA. 2.2. Taylor's expansion of the Laplace transform

In our set-valued framework, the Laplace transform can be expanded using either the probability mass function $p(\cdot)$, or the survival function $G^+(\cdot)$.

Expansion through $p(\cdot)$ By definition of the Laplace transform, we have:

$$\Psi_X(u) = \mathbb{E}\Big[\exp\big(-u_1X_1 - u_2X_2 - \cdots - u_nX_n\big)\Big] = \sum_{s \subset \{1,\ldots,n\}} p(s)e^{-u(s)}. \qquad \text{(eq. a.22)}$$

or equivalently, in terms of probability generating function[13]:

$$\mathbb{E}[z_1^{X_1} \cdots z_n^{X_n}] = \sum_{s \subset \{1,...,n\}} \Big[p(s) \prod_{i \in s} z_i\Big], \tag{eq. a.23}$$

with the convention $\prod_{i \in \emptyset} z_i = 1$. In other words, $p(s)$ is the coefficient of $\prod_{i \in s} z_i$ in the Taylor's expansion around $(0, 0, ..., 0)$ of the pgf.

Expansion through $G^+(\cdot)$ Alternatively, we can introduce $y_i = z_i - 1$, then equation (eq. a.23) becomes:

$$\begin{aligned}
\mathbb{E}[z_1^{X_1} \cdots z_n^{X_n}] &= \mathbb{E}[(1+y_1)^{X_1} \cdots (1+y_n)^{X_n}] \\
&= \sum_{s \subset \{1,...,n\}} \Big[p(s) \prod_{i \in s} (1+y_i)\Big] \\
&= \sum_{s, s^* \subset \{1,...,n\}} \Big[\sum_{s^* \subset s} p(s^*) \prod_{i \in s^*} y_i\Big] \\
&= \sum_{s, s^* \subset \{1,...,n\}} \Big\{\Big[\sum_{s^* \subset s} p(s^*)\Big] \prod_{i \in s^*} y_i\Big\} \\
&= \sum_{s^* \subset \{1,...,n\}} \Big[G^+(s^*) \prod_{i \in s^*} y_i\Big] \\
&= \sum_{s \subset \{1,...,n\}} \Big[G^+(s) \prod_{i \in s} y_i\Big].
\end{aligned} \tag{eq. a.24}$$

Therefore, $G^+(s)$ is the coefficient of the term $\prod_{i \in s} y_i$ in the Taylor expansion of the shifted pgf $\mathbb{E}[(1+y_1)^{X_1} \cdots (1+y_n)^{X_n}]$ around $(0, 0, ..., 0)$, or equivalently the Taylor's expansion of the pgf around $(1, 1, ..., 1)$.

As an illustration, let us consider the case $n = 2$. Then we have:

$$\mathbb{E}[z_1^{X_1} z_2^{X_2}] = p(\emptyset) + p(\{1\})z_1 + p(\{2\})z_2 + p(\{1,2\})z_1 z_2 \tag{eq. a.25}$$

$$\begin{aligned}
&= p(\emptyset) + p(\{1\})(z_1 - 1 + 1) + p(\{2\})(z_2 - 1 + 1) + p(\{1,2\})(z_1 - 1 + 1)(z_2 - 1 + 1) \\
&= 1 + [p(\{1\}) + p(\{1,2\})](z_1 - 1) + [p(\{2\}) + p(\{1,2\})](z_2 - 1) + p(\{1,2\})(z_1 - 1)(z_2 - 1) \\
&= G^+(\emptyset) + G^+((\{1\}))(z_1 - 1) + G^+((\{2\}))(z_2 - 1) + G^+(\{1,2\})(z_1 - 1)(z_2 - 1).
\end{aligned} \tag{eq. a.26}$$

Here, the expansion in (eq. a.25) is with respect to u_1 and u_2, whereas it is with respect to $u_1 - 1$ and $u_2 - 1$ in (eq. a.26).

[13]Since $(X_1, ..., X_n)$ has a discrete distribution, the probability generating function (pgf) exists and also characterizes its joint distribution.

OA. 2.3. Laplace transform in the LDPP framework

Let us now compute the Laplace transform, or equivalently the pgf of the set variable S, if S follows a LDPP. We have:

$$\begin{aligned}\mathbb{E}[z_1^{X_1}\cdots z_n^{X_n}] &= \sum_{s\subset\{1,\dots,n\}} \Big[G^+(s)\prod_{i\in s} y_i\Big] && \text{(eq. a.27)}\\ &= \sum_{s\subset\{1,\dots,n\}} \Big[\det K_s^+ \prod_{i\in s}(z_i-1)\Big]\\ &= \det\Big[Id + K^+\text{diag}(z_1-1, z_2-1, ..., z_n-1)\Big], && \text{(eq. a.28)}\end{aligned}$$

or equivalently,

$$\begin{aligned}\Psi_S(u) &= \mathbb{E}[\exp(-u(S))]\\ &= \det\Big[Id + K^+\text{diag}(e^{-u}-\mathbb{1})\Big], && \text{(eq. a.29)}\end{aligned}$$

where $e^{-u} = (e^{-u_1}, ...e^{-u_n})'$ and $\mathbb{1}$ is the vector with all elements equal to 1.

When the $u_1, ..., u_n$ are different, the term $u(S)$ can be interpreted as a total loss, with each u_i the individual loss given high risk. Then formula (eq. a.29) characterizes the distribution of the total loss.

OA. 2.4. Distribution of the size in the LDPP framework

By considering the special case $z_1 = \cdots = z_n = z$, we get in Section 2.3 the p.g.f. of the size $n(S)$ as:

$$\mathbb{E}[z^{n(S)}] = \prod_{i=1}^{n}[1+(z-1)\lambda_i] = \prod_{i=1}^{n}[1-\lambda_i+\lambda_i z].$$

Since $1-\lambda+\lambda z$ is the p.g.f. of the Bernoulli distribution $\mathcal{B}(1,\lambda)$, we deduce Proposition 12.

The computation above is valid for any symmetric matrix K^+ with eigenvalues smaller or equal to 1 in modulus. In particular, it remains valid, if some of the eigenvalues are equal to 1. Since these eigenvalues are $\lambda_i/(1+\lambda_i)$, where λ_i is an eigenvalue of Σ, this arises if some eigenvalues λ_i are infinite.

From the interpretation as sums of Bernoulli variables, we get the following result.

Proposition 24. *In the LDPP framework, the size $n(S)$ is constant if and only if the kernel K^+ has all eigenvalues equal to either 0, or 1. Then $n(S)$ is equal to the number of eigenvalues equal to 1.*

More generally, if K^+ has n_0 eigenvalues equal to 1, then $n(S)$ will be almost surely larger than n_0. This result is important to understand the behavior of the distribution of $n(S)$ in the illustrations of Section 6.2.

Online Appendix 3: Statistical inference

We will discuss statistical inference from i.i.d. observations $S_t, t = 1, ..., T$ of a random set.[14] We first consider unconstrained and constrained estimation. Then we develop the tests of the exchangeability and LDPP hypotheses.

OA 3.1 Estimation

When some results are well-known, we just recall the associated properties.

Appendix C.4.1 OA. 3.1.1. Unconstrained estimation

We have first to order the subsets by layer, i.e. by size, then within the layer. For a given ordering, the elementary probabilities $p(s)$, s varying, can be stacked in a vector $\boldsymbol{vec}(p)$ of dimension 2^n.

Proposition 25. *i) The maximum likelihood (ML) estimator of $p(s)$, s varying, is the sample frequency:*

$$\hat{p}_T(s) = \frac{\sum_{t=1}^{T} \mathbb{1}_{S_t=s}}{T}.$$

ii) Its asymptotic distribution for large T is:

$$\sqrt{T}\Big[\boldsymbol{vec}(\hat{p}_T) - \boldsymbol{vec}(p)\Big] \sim \mathcal{N}\Big(0, \boldsymbol{diag}[\boldsymbol{vec}(p)] - \boldsymbol{vec}(p)[\boldsymbol{vec}(p)]'\Big).$$

Note that the number of independent parameters is $2^n - 1$ and that the number T of observation has to be much larger than $2^n - 1$ for this asymptotic approximation to be valid. This explains the importance of constrained models.

OA 3.1.2. Exchangeable model

In exchangeable models, the sequence of observations can be summarized by the sequence of sizes $n(S_t)$, $t = 1, ..., T$, that is a sufficient statistic to estimate the distribution $P(\tilde{n}), \tilde{n} = 0, 1, ..., n$.

[14]Extensions to serially dependent observations S_t, $t = 1, ..., T$ is studied in Gouriéroux and Lu (2023b) under a Markov LDPP model and Gouriéroux and Lu (2023a) under a dynamic LDPP model with stochastic Wishart autoregressive kernel.

Proposition 26. *i) Under exchangeability, the maximum likelihood estimator of $P = (P(\tilde{n}))$ is the sample counterpart with:*

$$\hat{P}_T(\tilde{n}) = \frac{\sum_{t=1}^{T} \mathbb{1}_{n(S_t)=\tilde{n}}}{T}, \qquad \tilde{n} = 0, ..., n.$$

ii) Its asymptotic distribution for large T is:

$$\sqrt{T}(\hat{P}_T - P) \sim \mathcal{N}\Big(0, \boldsymbol{diag}(P) - PP'\Big).$$

OA 3.1.3. LDPP model

Let us denote $p(s, \Sigma) = \det \Sigma_s / \det(Id + \Sigma)$ the elementary probabilities. The log-likelihood function is:

$$\ell_T(\Sigma) = T \sum_s \hat{p}_T(s) \log p(s, \Sigma). \tag{5.1}$$

When T tends to infinity, $\frac{1}{T}\ell_T(\Sigma)$ tends to $\sum_s p_0(s) \log p(s, \Sigma) := c_\infty(\Sigma, \Sigma_0)$, where Σ_0 denotes the true value of matrix Σ and the LDPP model is assumed well-specified. This quantity is maximal if and only if $p(s, \Sigma) = p(s, \Sigma_0), \forall s$. This condition of observational equivalence can be written as "there exists a diagonal matrix D with diagonal elements $+1$ or -1 such that $\Sigma = D\Sigma_0 D$" [see Griffin and Tsatsomeros (2006)]. Therefore, Σ_0 is locally identifiable, but not globally identifiable. The local identification is sufficient for consistent ML estimation of matrix Σ.

The ML estimator of Σ is the solution:

$$\hat{\Sigma}_T = \arg\max_{\Sigma} \sum_s \hat{p}_T(s) \log p(s, \Sigma),$$

such that the nonnegativity condition $\Sigma \gg 0$ holds. The following proposition is proved in Gouriéroux and Lu (2023a).

Proposition 27. *In the LDPP framework, and standard regularity conditions*[15]

i) $\hat{\Sigma}_T$ is a consistent estimator of the true value Σ_0.

ii)

$$\sqrt{T}(\boldsymbol{vech}(\hat{\Sigma}_T) - \boldsymbol{vech}(\Sigma_0)) \sim \mathcal{N}(0, \Omega_0),$$

*where **vech** is the half-vectorization operator of a symmetric square matrix, and the asymp-*

[15] These conditions include the invertibility of the information matrix Ω^{-1}. In particular the matrix Σ cannot be block diagonal [Brunel et al. (2017)].

totic variance matrix Ω_0 is given by:

$$\Omega_0^{-1} = \sum_s \{ \frac{\det \Sigma_{0s}}{\det(Id+\Sigma_0)} H' J_s' (\Sigma_{0s}^{-1} \otimes \Sigma_{os}^{-1}) J_s H \} + H' [(Id+\Sigma_0)^{-1} \otimes (Id+\Sigma_0)^{-1}] H,$$

J_s and H being the selection matrices such that $vec\Sigma_s = J_s vec\Sigma$, $vec\Sigma = Hvech\Sigma$ and $\otimes$ denoting the Kronecker product.

Proof. We will perform a second-order expansion of the asymptotic log-likelihood to derive the information matrix and then the asymptotic variance-covariance matrix of the maximum likelihood estimator of parameter Σ_L. This expansion is performed in a neighborhood of the "true value" Σ_0, that is: $\Sigma = \Sigma_0 + \Delta$, where Δ is a small symmetric matrix.

The asymptotic log-likelihood function is :

$$L(\Sigma_0, \Delta) = \Sigma_s \frac{\det(\Sigma_{0s})}{\det(Id+\Sigma_0)} \log\det(\Sigma_{0s} + \Delta_s) - \log\det(Id + \Sigma_0 + \Delta),$$

where Δ is small. We have :

$$L(\Sigma_0, \Delta) = L(\Sigma_0, 0) + \Sigma_s \frac{\det(\Sigma_{0s})}{\det(Id+\Sigma_0)} \log\det(Id + \Sigma_{0s}^{-1}\Delta_s) - \log\det(Id + (Id+\Sigma_0)^{-1}\Delta).$$

Since : $\log\det(Id + A) =$Tr $\log(Id + A) \simeq Tr(A - \frac{1}{2}A^2)$, if A is small, we deduce :

$$\begin{aligned} L(\Sigma_0, \Delta) \quad &\simeq \quad L(\Sigma_0, 0) + \{\Sigma_s \tfrac{\det \Sigma_{0s}}{\det(Id+\Sigma_0)} Tr(\Sigma_{0s}^{-1}\Delta_s) - Tr[(Id+\Sigma_0)^{-1}\Delta] \\ &- \quad \tfrac{1}{2}\{\Sigma_s \tfrac{\det \Sigma_{0s}}{\det(Id+\Sigma_0)} Tr(\Sigma_{0s}^{-1}\Delta_s\Sigma_{0s}^{-1}\Delta_s) - Tr[(Id+\Sigma_0)^{-1}\Delta(Id+\Sigma_0)^{-1}\Delta]\}. \end{aligned}$$

This differential expression is written in term of matrix Δ. Then the last term has to be rewritten in terms of vec or vech operators.

Since : $Tr(AX'BXC) = (vecX)'(CA \otimes B')vecX$, where $\otimes$ denotes the Kronecker product [see Henderson and Searle (1979), eq . 9], we can typically write twice the second-order term of the expansion as :

$$-\left\{ \frac{\det \Sigma_{0s}}{\det(Id+\Sigma_0)} (vec\Delta_s)'(\Sigma_{0s}^{-1} \otimes \Sigma_{0s}^{-1})(vec\Delta_s) \right\} - (vec\Delta)'[(Id+\Sigma_0)^{-1} \otimes (Id+\Sigma_0)^{-1}]vec\Delta.$$

Let us denote J_s the selection matrix of dimension $(n(s), n)$ such that $vec\Delta_s = J_s vec\Delta$ and the matrix H such that $vec\Delta = Hvech\Delta$, we deduce the closed form expression of the information matrix, that is the inverse of the asymptotic variance-covariance matrix as :

$$V^{-1} = \Sigma_s \left\{ \frac{\det \Sigma_{os}}{\det(Id+\Sigma_0)} H' J_s' (\Sigma_{0s}^{-1} \otimes \Sigma_{0s}^{-1}) J_s H \right\} + H'[(Id+\Sigma_0)^{-1} \otimes (Id+\Sigma_0)^{-1}]H.$$

□

In the LDPP framework, some of the parameters can also be estimated by using a method of moments. Indeed, for $i \neq j$:

$$p(\{i\})/p(\emptyset) = \sigma_{ii}, \qquad p(\{i,j\})/p(\emptyset) = \sigma_{ii}\sigma_{jj} - \sigma_{ij}^2.$$

Thus, by focusing on subsets of size 1 or 2, we can identify all the diagonal elements of Σ, as well as all the off-diagonal elements of Σ, up to their sign.

As a consequence, by replacing $p(\{i\}), p(\emptyset), p(\{j\})$ and $p(\{i,j\})$ by their empirical counterparts, we obtain moment estimates of σ_{ii} and $|\sigma_{ij}|$. This method can be used to obtain a first step estimate, which can be used as the initial value for the algorithm of likelihood maximization. However, this method of moment has two weaknesses. First, this method of moment does not identify the signs of the off-diagonal terms. Second, it is generically less efficient than the maximum likelihood approach.

OA 3.2. Test of constrained models

The approach is standard and based on likelihood ratios. We denote $\mathcal{H}_1$ the unconstrained model, $\mathcal{H}_{OE}$ the null hypothesis of exchangeability, $\mathcal{H}_{OD}$ the null hypothesis of LDPP model, and $\hat{\ell}_T, \hat{\ell}_T^E, \hat{\ell}_T^D$ the associated estimated log-likelihoods.

Proposition 28. *Asymptotically valid likelihood ratio tests at level α are:*

i) for exchangeability:

$$\text{if} \qquad 2T(\hat{\ell}_T - \hat{\ell}_T^E) < \chi^2_\alpha(2^n - n - 1), \quad \text{accept } \mathcal{H}_{OE}, \text{reject it, otherwise.}$$

ii) for LDPP:

$$\text{if} \qquad 2T(\hat{\ell}_T - \hat{\ell}_T^D) < \chi^2_\alpha(2^n - 1 - \frac{n(n+1)}{2}), \quad \text{accept } \mathcal{H}_{OD}, \text{reject it, otherwise.}$$