Chapitre 1

Literature

Recent literature analysed the psychological phenomenon of correlation neglect (Eyster and Weizsäcker, 2011; Enke and Zimmermann, 2015; Levy and Razin, 2015; Ortoleva and Snowberg, 2015). Correlation neglect implies that individuals underappreciate the correlation between state variables or different events they observe. This cognitive bias has negative spillovers for individuals decisions making and lead to overconfidence in market settings which predicts bubbles and crashes. A small set of previous experiments have examined individuals' responses to correlations in informational sources and have found that subjects have limited attention and find it cognitively challenging to work with joint distributions of random variables (Eyster and Weizsäcker, 2011). However, in this literature, there is a conceptual difference in the meaning of correlation neglect (structure of the environment and empirical analysis of historical data). The literature on boundedly rational and beliefs formation in networks uses the structure of the information and analyses a double-counting problem in the informations sources when people update their beliefs about a state variable (Enke and Zimmermann, 2015). As result, subjects in experiments on group communication (DeMarzo et al., 2003) or political competition and voting behavior (Ortoleva and Snowberg, 2015) overweight the impact of informational redundancies in their beliefs.

Enke and Zimmermann (2015) used the structure of decision problem and analysed double counting problem in beliefs formation. They find that experimental subjects in a relative simple setting neglect correlations in information sources when forming beliefs with heterogeneity at the individual level. They suggest a measure of individual correlation neglect in the between subjects design under the assumption that signals are drawn from a truncated discretized normal distribution with mean μ - the true value of the state - and standard deviation $\sigma = \frac{\mu}{2}$. Truncation implies that signals belong to the interval $[0, 2\mu]$ in order to avoid negative signals and then negative correlations. But in real life, people face many situations involving negative and positive informations (signals) about events. Informations might be positively or negatively correlated. In their experiment, two computers, A and B, generate two *"iid"* unbiased signals s_A and s_B with ($s_h \sim N(\mu, (\frac{\mu}{2})^2), h \in \{A, B\}$). Subjects observe these signals as numbers that they must use to estimate the number of items in an imaginary container. In the correlated treatment, subjects observe the realizations of a computer A (s_A) and the mean realization of two computers $\tilde{s}_B = \dfrac{s_A + s_B}{2}$ so that the two signals are correlated with a correlation of 71% [1]. Subjects in the control condition observe two independent signals (s_A and s_B). For rational estimation of the number of item, subjects in the correlated treatment must take into account the information about the correlation of two signals given the structure of the environment when the unbiased estimate of the number of items is the empirical mean of two signals in control condition. Therefore, a rational subject must extract s_B from $\tilde{s}_B$ and compute the mean of s_A and s_B as an estimate of the number of items in the container. The following rule is used for each subject in the correlated treatment when he tries to extract the right signal s_B :

$$\hat{s}_B = \chi \tilde{s}_B + (1 - \chi) s_B$$

χ is an individual measure of correlation neglect that captures subjects ability to extract the right signal into $\hat{s}_B$ when authors consider only the structure of the environment.

$$\begin{cases} \chi = 0 & \text{for rational subject} \\ \chi = 1 & \text{for full correlation neglecter} \end{cases}$$

As result, people's beliefs in the correlated treatment deviate from rationality because subjects neglect informational redundancies. Their individual measure of correlation neglect can be apply in differents settings on informational structure (e.g *Group communication*, *Voting Behavior, Political competitions, etc.*). While a part of literature focuses on the structure of the environment, the other analyses people's limited attention on joint distribution of random variables when engaging in empirical analysis of historical and empirical data (Kallir and Sonsino, 2009; Eyster and Weizsäcker, 2011).

Kallir and Sonsino (2009) find that changing the correlation of a portfolio-choice problem leads to little or no change in participants decision making. In their experiment, subjects observed historical data on the joint distribution of the realized returns of two virtual assets with different levels of correlation for 12 preceding periods ; and they have to predict the realized returns of the first asset in four additional observations when observing the returns of the second asset. In this predictions-allocations problems, the results show that subjects recognize shifts in correlation in their prediction tasks but fail to account for this correlation in their allocations decisions. Therefore, correlation neglect predicts no change in participants' behavior. They shed additional light on the cognitive nature of the bias that is consistent with

1. See appendix A

the interpretation of correlation neglect as deriving from limited attention. No formal measure of individual correlation neglect has been suggested.

Eyster and Weizsäcker (2011) focus on the impact of correlation neglect on financial decision making because an investor who fails to account for the correlation when allocating his financial portfolio can hold a portfolio that contains undesirable risks. They suggest a measure of individual correlation neglect in a series of controlled experiments using a framing variation in which each participant faces two versions of the same portfolio-choice problem. The assets in the correlated frame are linear combinations of those in the uncorrelated frame and span exactly the same set of earnings distributions. Under the hypothesis that people correctly perceive the covariance structure, the framing variations does not affect behavior. By ensuring that participants understand the payoff structure and the co-movements of the assets returns, they find that behaviors change strongly. People ignore the correlation and treat correlated assets as independent following sometimes a simple "1/N heuristic" which is investing equal shares of financial portfolio into all available assets. They measure people "ignorance" using the following transformation of the matrix of variance-covariance with penalties on variance and covariance terms.

$$V = \begin{pmatrix} (\sigma_1^2)^l & k.sgn(\sigma_{12})|\sigma_{12}|^l \\ k.sgn(\sigma_{12})|\sigma_{12}|^l & (\sigma_2^2)^l \end{pmatrix}$$

k and l represent the parameters of correlation neglect and variance neglect (respectively) and are estimated for each subject in the experiment. Then, they classify subjects in 3 differents groups according the severity of correlation neglect.

This conceptual difference according the context and the structure of the decision problem and various experimental approaches (*within and between-subjects designs*) raises the difficulty to apply a measure of one paper to evaluate correlation neglect in another. To the best of our knowledge, there exists no a single measure of correlation neglect that can be applied in different contexts.

Our paper is related to this literature on correlation neglect and we propose a measure that can be used as general measure of correlation neglect regardless of contexts. We need not to assume any hypothesis about individuals' preferences, but only assessing their beliefs about the distribution of state variables in our experiment. Then, we compute a measure of individual correlation neglect.

Chapitre 2

Experiment

A decision maker can identify correlation between state variables either by engaging in empirical analysis of historical data or by analyzing the structure of the environment. Correlation neglect is present, when information, that can help to identify the correlation, is available but is ignored in the decision process. It is important to distinguish between the two dimensions because it allows to localise the cognitive shortcoming of judgement and because it might affect the choice of policy if one wants to mitigate the ignorance.

Following this reasoning, we propose to measure correlation neglect in two ways : first, the correlation of state variables is presented by the structure of the decision problem (e.g Eyster and Weizsäcker (2011), Enke and Zimmermann (2015)), or second, by observing realizations of both variables (e.g Kallir and Sonsino (2009)).

As in the literature on correlation neglect, we presume that ignoring the correlation between two random variables affects beliefs about the joint distribution of those variables. However, different to the existing literature, we propose to measure correlation neglect directly by eliciting (subjective) beliefs and not indirectly via observed choices.

2.1 Experimental Design

The basic set-up of our experiment consists of two urns, **Urn 1** and **Urn 2**, containing N_1 and N_2 balls, respectively. Balls are either blue B or green G with $B_1 + G_1 = N_1$ and $B_2 + G_2 = N_2$ and the distribution is represented by the ratio of blue balls, $b_1 = B_1/N_1$ and $b_2 = B_2/N_2$. B_1 and G_1 are respectively the number of blue and green balls in **Urn 1** while B_2 and G_2 represent the number of blue and green balls in **Urn 2**. Then a number of balls D_1 are drawn from **Urn 1** without replacement and placed in **Urn 2**, from which then D_2 balls are drawn again without replacement. This procedure is repeated S times, each time resetting both urns to the original set-up. The task of the participant is to give a personal evaluation of the following three distributions : distribution of variable X : representing the distribution of blue balls

(X_B) or green (X_G) in D_1 over S, the distribution of variable Y, representing the distribution of blue (Y_B) or green (Y_G) balls in D_2 over S, and the distribution of variable Z, representing their joint distribution over S.

	Simple Set Up												
no	**Urn 1**				**Urn 2**				$E[X]$	$E[Y]$	$E[XY]$	Cov	$Corr$
	N_1	b_1	D_1	$\frac{D_1}{N_1}$	$\mathbf{N_2}$	b_2	D_2	$\frac{\mathbf{D_2}}{(\mathbf{N_2+D_1})}$					ρ
									$X = X_B, Y = Y_B$				
1	2	0.5	1	0.5	**2**	0.5	1	**0.33**	0.5	0.5	**0.33**	**0.08**	**0.33**
2	2	0.5	1	0.5	**4**	0.5	1	**0.20**	0.5	0.5	**0.30**	**0.05**	**0.20**
3	2	0.5	1	0.5	**8**	0.5	1	**0.11**	0.5	0.5	**0.28**	**0.03**	**0.11**
4	2	0.5	1	0.5	**200**	0.5	1	**0.005**	0.5	0.5	**0.25**	**0.005**	**0.005**
									$X = X_B, Y = Y_G$				
5	2	0.5	1	0.5	**2**	0.5	1	**0.33**	0.5	0.5	**0.16**	**-0.08**	**-0.33**
6	2	0.5	1	0.5	**4**	0.5	1	**0.20**	0.5	0.5	**0.20**	**-0.05**	**-0.20**
7	2	0.5	1	0.5	**8**	0.5	1	**0.11**	0.5	0.5	**0.22**	**-0.03**	**-0.11**
8	2	0.5	1	0.5	**200**	0.5	1	**0.005**	0.5	0.5	**0.245**	**-0.005**	**-0.005**

TABLE 2.1 – Experimental Design : Variation of experimental parameters and correlation coefficients.

A simple case of our experiment is shown by the following example that is summarized in Table 2.1. The urns contain 2 balls, one blue and one green, each. One ball is drawn from each urn, with a total of S=100 repetitions. x_B is the number of times out of 100 repetitions a blue ball would be drawn first, y_B is the number of times out of 100 a blue ball would be drawn second and z_B is the number of times out of 100 where both draws would be blue. The corresponding questions eliciting the distribution of those variables are :

1. *"What are the chances out of 100 that a blue ball is drawn from the first urn ?"*

2. *"What are the chances out of 100 that a blue ball is drawn from the second urn ?"*

3. *"What are the chances out of 100 that a blue ball is drawn from both urns ?"*

With the response "x_B *out of 100*," question 1 elicits $E[X] = \Pr[X = B] = \Pr[D_1 = B] = x_B/100 = \bar{x}_B$. Response to question 2 reveals $E[Y] = \Pr[Y = B] = \Pr[D_2 = B] = y_B/100 = \bar{y}_B$ and to question 3, $E[Z] = E[XY] = \Pr[X = B, Y = B] = \Pr[D_1 = B, D_2 = B] = z_B/100 = \bar{z}_B$. The correlation is obtained simply by $\rho = (\bar{z}_B - \bar{x}_B\bar{y}_B)/\sqrt{\bar{x}_B\bar{y}_B(1-\bar{x}_B)(1-\bar{y}_B)}$. Given the experimental design of the simple example, the theoretical correlation between the two random variables is 0.33.

With this structure, we introduced a correlation between X and Y when taking a ball from the first urn and put it in the second urn. Because ignoring this correlation affects beliefs about the joint distribution of X and Y, Table 2.2 presents beliefs' prediction for rational subjects and full correlation neglecters for 4 parameterizations presented in Table 2.1. Someone who fully neglects the structure of correlation thinks that there is no link between X and Y and then :

$$P[X = x_i, Y = y_j] = P[X = x_i] \times P[Y = y_j] \text{ for } i, j \in \{Blue, Green\}$$

	P[X=1, Y=1]	
Treatment correlation	Rational beliefs	Full Corr. neglect beliefs
0,33	0,33	0,25
0,2	0,3	0,25
0,11	0,27	0,25
0,005	0,25	0,25

TABLE 2.2 – beliefs prediction for joint distribution

Our design is structured so that for all beliefs elicitations problem $E[X] = E[Y] = \frac{1}{2}$. The *Simple Set Up* allows the correlations to lay between -0.33 and 0.33[1]. In the *Simple Set Up* we restrain $N_1 = 2, b_1 = b_2 = 0.5$ and $D_1 = D_2 = 1$. By varying N_2, the size of the second urn, we can manipulate the level of correlation to be between 0 and 0.33[2]. And by varying whether the subjective expectation for the second draw concern the same color as the one in the first draw or the other, we manipulate the direction of the correlation to be positive or negative. Rows (1) - (8) of Table 2.1 shows 8 possible parameterizations resulting in correlations of $\{-0.33, -0.20, -0.10, -0.005, 0.005, 0.10, 0.20, 0.33\}$ covering uniformly the range of possible values.

This first design allows to measure people understanting of the correlation when they face situations which introduce the correlation by taking one variable as its combination with the other one. For instance, Eyster and Weizsäcker (2011) construct portfolio choice problems with state-dependent returns using framing variation in which each participant faces two versions of the same portfolio-choice problem. Across the two framing variations, they switch asset correlation on and off as presented in table 2.3

	State-dependent returns	
	{X(1), X(2), X(3), X(4)}	{Y(1), Y(2), Y(3), Y(4)}
portfolio 1	A = {12, 24, 12, 24}	B = {12, 12, 24, 24}
portfolio 2	C = {12, 24, 12, 24}	D = {12, 18, 18, 24}

TABLE 2.3 – Structure of portfolio choice problem.

In portfolio 1 there is no correlation between asset A and B. Portfolio 2 is constructed such that the returns of $C = A$ and $D = \dfrac{A+B}{2}$, thus introducing the correlation between C

1. By switching the color of the ball drawn in the second urn we allow correlation to be positive or negative
2.

$$\lim_{N_2 \to \infty} \rho_{X,Y} = 0$$

(proof in appendix A.2)

and D. Under the hypothesis that people correctly perceive the correlation structure, this framing variation does not affect their behaviour. Our simple design allows to measure people understanding of the correlation in this kind of situations before making their decisions.

Instead of observing the structure of the environment (decision problem), subjects may face situations in which they observe historical data of state variables. This situation is illustrated in Kallir and Sonsino (2009) where subjects observe the joint distribution of realized returns of two virtual assets with two levels of returns (high and low) for 12 preceding periods. They consider five predictions problems involving five different levels of correlation between assets returns and subjects are requested to predict returns for 4 additional periods under the assumption that future returns are sampled from the empirical distribution.

We integrate this situation in our experiment and subjects observe S draws from the first and the second urn simultaneously. We allow the number of draws S to be endogenous to each participant in the experiment. Then, each subject can decide on the number of draws that he wants to observe. By endogenizing S, the number of draws, participants control the quantity of information that they have, a possible source of debiasing. The task of the subject is to give his personal evaluation of following distributions : distribution of variable X : representing the distribution of blue (X_B) or green (X_G) balls in D_1 over 100, the distribution of variable Y, representing the distribution of blue (Y_B) or green (Y_G) balls in D_2 over 100, and the distribution of variable Z, representing their joint distribution over 100. The corresponding questions eliciting the distribution of those variables are :

1. *"In how many out of 100 draws do you think that* ***a blue ball or a green ball*** *is drawn from the first urn ?"*
2. *"In how many out of 100 draws do you think that* ***a blue ball or a green ball*** *is drawn from the second urn ?"*
3. *"In how many out of 100 draws do you think that* ***a blue ball*** *is drawn from the first urn and* ***a green ball*** *from the second urn ?"*
4. *"In how many out of 100 draws do you think that* ***a green ball*** *is drawn from the first urn and* ***a blue ball*** *from the second urn ?"*

2.2 Treatments

We consider three different presentations. First, a **presentation of the structure** of the situation, but no demonstration of realizations, i.e., $S = 0$ and questions as in the structural presentation above. Second, no information on the structure, but a **time series** showing actual realizations of a certain amount S of draws and questions as in the empirical presentation. Third, no information on the structure, but a **time series** showing joint distributions of variables as relative frequencies in matrix form and questions as in the empirical presentation.

Chapitre 3

Empirical Measure

In this section, we present our measure of individual correlation neglect and the others measures in the literature.

3.1 Simple Set Up

In this paper, we presume correlation neglect to be an individual trait of a person and we propose a measure of this caracteristic. To compute a subjective measure of ***"individual correlation"*** when asking for their beliefs about the joint distribution of state variables in the **Empirical treatment**, we use a measure of correlation for bivariate data, the so called ***"Phi Coefficient."*** This is one of the straightforward and usefull methods to assess the correlation between two bivariate variables and it has the same interpretations as pearson's correlation. The *"Individual Phi Coefficient"* for each subject is compared to the true value of the correlation allowed by our experiment and in the same treatment.
During the experiment and in empirical treatments, subjects answer differents questions about bivariate variables and for each subject, we construct a 2×2 matrix corresponding to his answers. Let subject *"i"* when answer to questions in treatment *"j"* forms the following *2 × 2 Matrix.*

		Urn 2 : Variable Y		
		Blue	Green	Total
Urn 1 : Variable X	Blue	a_i^j	b_i^j	e_i^j
	Green	c_i^j	d_i^j	f_i^j
	Total	g_i^j	h_i^j	$n = 100$

In the experiment, e_i^j and f_i^j represent the subjectives distributions of variable X for individual i. g_i^j and h_i^j represent the distribution of variable Y (representing the color of the ball drawn in the second urn) in the same treatment. With this presentation, the value of correlation for individual i in treatment j, is computed as follow :

$$\phi_i^j = \frac{a_i^j \times d_i^j - c_i^j \times b_i^j}{\sqrt{(e_i^j \times f_i^j \times g_i^j \times h_i^j)}}$$

Since ϕ_i^j is a subjective value of the correlation for individual i in the treatment j, this is compared to the right value of correlation in the same treatment (called ϕ^j). Our measure is defined as follow :

$$\chi_i^j = \phi^j - \phi_i^j$$

χ_i^j quantifies individual i *correlation neglect* in treatment j. This framework allows χ_i^j to belong in the interval $[-2, 2]$ where near to 0 represents rational subjects.

In the *structurals treatments* we don't need to use the definition of *"Phi Coefficient;"* subjects responses are used to compute their subjective correlation in the corresponding treatment using the formula :

$$\rho_i = \frac{cov_i(X,Y)}{\sigma_{i,X}\sigma_{i,Y}} = \frac{P[X=x, Y=y] - E[X]E[Y]}{\sqrt{P[X=x](1-P[X=x]) \times P[Y=y](1-P[Y=y])}}$$

This value is then compared to the theoretical value of correlation as describe above to quantify their neglect.

For the purpose of some statistical analysis, because our beliefs formation tasks allow for 3 differents presentations of the information, we compute a single measure of individual correlation neglect (median correlation neglect) for each subject and each informational presentation. We compute this measure by taking a median of j correlation neglect parameters at the same informational presentation (k) :

$$\chi_i^k = med(\phi^{j,k} - \phi_i^{j,k})$$

Then, each subject has one value of correlation neglect parameter by type of informational presentation.