

# Combining Forecasts: Why Decision Makers Neglect Correlation

Gilat Levy and Ronny Razin, LSE<sup>1</sup>

*Abstract:* We suggest a framework to analyse how individuals combine multiple sources of information to form predictions. We assume that individuals understand each source of information separately, but they are not certain about the correlation between them. We bound the ability of individuals to grasp such correlations by a notion of “correlation capacity”. We show that given this capacity there is a set of predictions which is completely characterised by two parameters: the correlation capacity parameter and the naive combination of forecasts which ignores correlation. The analysis yields two countervailing effects on behaviour. A higher correlation capacity creates more uncertainty and therefore possibly conservative behaviour. On the other hand, when the naive combination is relatively precise, it can induce risky behaviour. We show how this trade-off affects behaviour in different applications, including financial investments and CDO ratings. Specifically we show that complex assets are likely to lead to complete neglect of correlation by individuals.

*“...If we assumed that states had the same overall error as in the FiveThirtyEight polls-only model but that the error in each state was independent, Clinton’s chances would be 99.8 percent, and Trump’s chances just 0.2 percent. So assumptions about the correlation between states make a huge difference. Most other models also assume that state-by-state outcomes are correlated to some degree, but based on their probability distributions, FiveThirtyEight’s seem to be more emphatic about this assumption,”* Nate Silver, FiveThirtyEight discussing his predictions about the 2016 US Presidential election.

## 1 Introduction

When confronted with multiple sources of information, we often have a better understanding of each information source separately than we do of how the sources relate to one another. This tendency is apparent in many situations, both when individuals, even experts, analyze data, or when organisations make predictions. In the finance literature this has been long

---

<sup>1</sup>We thank Eddie Dekel, Andrew Ellis, Jeff Ely, Erik Eyster, Daniel Krahmer, Francesco Nava, Alp Simsek and Joel Sobel for helpful comments. We also thank seminar participants at the Queen Mary Theory workshop, Bocconi, PSE, ESSET 2015, Manchester University, University of Bonn, LBS, Ecole Polytechnique, Bristol University, St. Andrews, Rotterdam, Zurich, York-RES Game Theory symposium, Warwick/Princeton Political Economy workshop, King’s College and Queen Mary Political Economy workshop, and the UCL-LSE Theory workshop. ERC grant SEC-C413 provided valuable financial support for this research.

recognised.<sup>2</sup> Jiang and Tian (2016) observe that “It is well known that the correlation estimation is more difficult than the estimation of the expected mean or volatility from both statistical perspective and econometric perspective...”. They point to several problems in estimating the correlation or covariance process, including the lack of enough market data, limitation in estimation methodology, instabilities in the correlation process and the increasingly interconnected market patterns. In practice, this is also a well known issue. The US financial crisis inquiry (FCIC) report from 2011 cites the rating agency Moody’s acknowledgment that “In the absence of meaningful default data, it is impossible to develop empirical default correlation measures based on actual observations of defaults.”

In the face of such uncertainty about correlation between sources of information, treating information as independent is often used as a benchmark around which different scenarios are then considered. For example the FCIC report documents how rating agencies have followed a particular procedure by which CDOs were rated. Analysts (“quants”) used the individual level data of the loans making up a CDO, and then considered different correlation scenarios: “The M3 Prime model let Moody’s automate more of the process...Relying on loan-to-value ratios, borrower credit scores, originator quality, and loan terms and other information, the model simulated the performance of each loan in 1250 scenarios.”

When committees chose their final ratings given these scenarios, individual preferences and the culture of the organisation played an important part. Anil K Kashyap’s paper prepared for the FCIC states: “There is an inherent tendency for the optimists about the products to push aside the more cautious within the organization”.<sup>3</sup> Other organizations (and certainly many more after the 2008 crisis) became more cautious due to this uncertainty. Arnold Cattani, chairman of Mission Bank in Bakersfield, described deciding to sell the bank’s holdings of mortgage-backed securities:.. “I told him [the CFO] to sell them, sell all of them, then, because we didn’t understand it, and I don’t know that we had the capability to understand the financial complexities; didn’t want any part of it.”

The recent attempts of the polling industry to predict the results of the 2016 US Presidential election followed a similar pattern. Poll aggregators produced predictions that were all based on the same survey data from the different US states. The perception on the eve of the election that a win for Donald Trump is unlikely was often motivated by the low probability that he might win in a combination of states, such as Pennsylvania, Michigan and North Carolina. The low probability given to this event stems from treating the different state polls as independent.

On the other hand, Nate Silver’s FiveThirtyEight provided one of the most cautious predictions for Hillary Clinton winning. Nate Silver outlined the different assumptions underlying his model, “... Assumption No. 4: State outcomes are highly correlated with one another, so polling errors in one state are likely to be replicated in other, similar states...If we assumed

---

<sup>2</sup>This is the motivation behind papers such as Duffie et al (2009).

<sup>3</sup>[http://faculty.chicagobooth.edu/anil.kashyap/research/papers/lesson\\_for\\_fcic.pdf](http://faculty.chicagobooth.edu/anil.kashyap/research/papers/lesson_for_fcic.pdf)

that states had the same overall error as in the FiveThirtyEight polls-only model but that the error in each state was independent, Clinton’s chances would be 99.8 percent, and Trump’s chances just 0.2 percent. So assumptions about the correlation between states make a huge difference.”

Similar patterns of behaviour also arise when non-experts are asked to combine forecasts. Maines (1996) finds that when combining multiple sources individuals often take into account relative historical accuracy of individual forecasters but not necessarily the relative historical dependence between them. Moreover, individuals differ in the degree to which they do take this into account. These experiments also point to particular patterns in forming predictions. Maines (1996) documents conservative behaviour in the face of multiple forecasts. Budescu and Yu (2007) show that decision makers become more confident when the advisors’ estimates were more extreme, advisors were in higher agreement with each other, and when advisors’ original signals were uncorrelated.

In this paper we suggest a framework to model how individuals combine forecasts in complex environments. We consider an environment in which an agent observes forecasts about a potentially multidimensional state of the world,  $\omega \in \Omega^n$ . For each element  $\omega_i$  in  $\omega$ , the agent observes possibly multiple forecasts, each a probability distribution over  $\Omega$ . To combine the multiple forecasts into a prediction about  $\omega$ , the agent considers a set of possible joint information structures that could have possibly yielded these forecasts. For each joint information structure in this set, that is consistent with the multiple forecasts, the agent derives a Bayesian prediction over the state of the world. This process yields a set of predictions about  $\omega$  that is the focus of our analysis.

Our model allows for two sources of correlation in the consideration set of the agent. First, the elements in  $\omega$  may be correlated; for example, a poll aggregator may base its prediction on the assumption that shocks to voters’ preferences across US states might be correlated, or a rating agency might assume that there is correlation in the default probabilities of the loans comprising a CDO. Another source of correlation might come from the statistical relation between the different observed forecasts. In the polling example, these might be due to biases in polling techniques. For the CDOs, badly designed stress tests that are applied across the board.<sup>4</sup>

Our key modelling contribution is to introduce the notion of “correlation capacity” to restrict the set of joint information structures and priors the agent considers. We model this capacity by a parameter which bounds the degree of *pointwise mutual information (PMI)* between information sources. PMI relates to the “distance” between the joint distribution

---

<sup>4</sup>For CDOs, Duffie et al (2009) suggest: “An example of an important factor that was not included in most mortgage portfolio default loss models is the degree to which borrowers and mortgage brokers provided proper documentation of borrowers’ credit qualities. With hindsight, more teams responsible for designing, rating, intermediating, and investing in subprime CDOs might have done better by allowing for the possibility that the difference between actual and documented credit qualities would turn out to be much higher than expected, or much lower than expected, in a manner that is correlated across the pool of borrowers.”

and the independent benchmark, that is, the multiplication of the marginal distributions. The higher is the capacity, the more correlation levels can be considered. Modelling the perceptions of individuals about correlation in this way is general and encompasses the families of distributions and copulas considered in the literature.

The capacity for correlation can be interpreted as a characteristic of an individual or of an organizational strategy. The capacity can also change following shocks or outside pressures. For example, for the case of rating agencies, the FCIC reports that “Moody’s officials told the FCIC they recognized that stress scenarios were not sufficiently severe, so they applied additional weight to the most stressful scenario.” And moreover “The output was manually “calibrated” to be more conservative...analysts took the “single worst case” from the M3 Subprime model simulations and multiplied it by a factor in order to add deterioration.”

When individuals have no capacity for correlation, they only consider joint information structures that are conditionally independent. In this case there is a unique rational belief arising from combining forecasts, which is proportional to a simple multiplication of the forecasts. We denote this as the Naive-Bayes (NB) belief.

Our main result is to characterise the set of predictions that arise from the agent’s consideration set when she has capacity for correlation. We show that the set of predictions can be fully characterised by two sufficient statistics: the capacity for correlation and the NB belief. This is a useful result as these two statistics can be computed from observable data.<sup>5</sup> To follow the CDO example above, our model provides a rationalisation to the procedural behaviour of companies like Moody’s and to the patterns of behaviour outlined above: the set of predictions is generated using individual loan data and a fixed set of scenarios.

We show that the set of predictions is convex and compact, and is monotonic (set-wise) in the correlation capacity. Therefore, the model provides a nuanced link between complexity and confidence. Individuals or organizations that have high correlation capacity will perceive more complex environments and will end up with a larger set of predictions. In this sense, correlation capacity can serve as a foundation for models of ambiguity. One can interpret the set of predictions as a set of priors of an individual who has ambiguity. The convexity of the set allows one to use many of the existing models of ambiguity to yield predictions about behaviour, as we do in one of the applications below. A larger correlation capacity will lead then to more ambiguity aversion, and potentially, more cautious behaviour as a result. We term this the *cautiousness effect*.

Finally we show that if the NB belief is close to being degenerate, the set of predictions shrinks and converges to the NB belief. Our characterisation therefore provides a foundation for correlation neglect. In the CDO example, if the naive interpretation of individual loan data, the NB belief, points to low risk of default of the CDO as a whole, the set of predictions

---

<sup>5</sup>The Naïve-Bayes benchmark can be computed from the observed forecasts, with no knowledge of the particulars of information structures needed. Correlation capacity can be extracted from individual choice data as we explain in Section 4.1.

converges to the unique belief arising from correlation neglect. This would imply that even a heterogenous committee of decision makers, faced with the same scenarios, will all take decisions as if they neglect correlation. We term this the *Naive-Bayes effect*.

We analyze two applications to highlight the implications of the above results. The first is a simple investment model, in which individuals can invest in a risky or safe asset and have ambiguity aversion. Each investor receives information about the asset but these signals can be correlated. In the first period individuals invest and can then adjust their investments in the second period following the observations of others' behaviour. We show that when the number of investors is low, then investors with a high correlation capacity will reduce their risky investment in the second period. However, when the number of investors is large, the NB belief might become more precise. As a result, substantial cautious and risky shifts may occur in large groups.<sup>6</sup> We therefore identify a non-monotonic effect of the number of investors on second period behaviour.

We use this to interpret the behaviour before and after the 2008 financial crisis, illustrating that markets can move quickly from boom to bust. Pre-2008, as the FCIC report illustrates, investors did not appreciate the level of correlation across assets. Post-2008, investors shied away from risk as they had realised that they have little knowledge of the true model of correlation across assets. Our model can explain both these phenomena by using the two effects we uncover. First, the capacity for correlation might have increased post 2008, contributing to a shift from more risky to more cautious behaviour. Second, the volume of trade in a market can be linked to the precision of the NB belief, where a low volume of trade will indicate a weaker belief. Thus both high correlation capacity and a low post-2008 trade volumes will contribute to low confidence and more cautious behaviour.

Our second application is a simple model of the rating of a complex financial assets such as CDOs. In this application we focus on the correlation across individual loans (that are the elements of the state of the world, i.e., the default status of the CDO). Typically, a CDO is rated as triple A if the probability of default for at least some share  $\alpha$  of its loans is low enough. This means that one needs to assess a combination of states rather than the occurrence of one specific state. We show that our analysis extends to this environment as well. When the CDO includes sufficiently many loans, and when  $\alpha$  is sufficiently high, the high rating will be awarded even when rating agencies consider many correlations or when they focus on the worst case scenario. Therefore, this application highlights a connection between complexity (modelled as the number of individual elements in a financial asset) and correlation neglect. The more complex the assets, the more our agents will analyze them as if they completely neglect the possible correlations between the individual elements of the asset.

---

<sup>6</sup>When the different forecasts are overall pessimistic, both effects, cautiousness and the Naïve-Bayes, will point in the same direction, towards more cautious behaviour. When the forecasts are overall optimistic on the other hand, there maybe a tension between the two effects, determining whether behaviour will become more risky or more cautious.

A recent literature has studied correlation neglect, i.e., a behavioral assumption that individuals neglect taking account of possible correlation between multiple sources of information. For example DeMarzo et al (2003) and Glaeser and Sunstein (2009) study how this affects individual beliefs in groups, Ortoleva and Snowberg (2015) study its implications for individual political beliefs and Levy and Razin (2015a, 2015b) focus on the implication of correlation neglect in voting contexts. Enke and Zimmerman (2013), Kallir and Sonsino (2009) and Eyster and Weiszacker (2011) show how correlation neglect arises in experiments, with the latter two focusing on financial decision making. Alternatively, Ellis and Piccione (forthcoming) use an axiomatic approach to represent decision makers affected by the complexity of correlations among the consequences of feasible actions.

Our key result indicates that in some environments correlation neglect will arise endogenously. While we do not provide behavioural foundations for correlation capacity, our analysis is general and the set of beliefs we derive will nest any model that allows for a particular information structure or assumptions on beliefs over correlation. We show that the set of rationalisable beliefs shrinks if the information contained in the Naive-Bayes benchmark that accords with full correlation neglect- is precise enough. Thus, in political environments with many polls, markets with many investors, or for securities which are composed of many assets, correlation neglect is likely to arise.

Since the 2008 financial crisis, the issue of the uncertainty about correlation in default rates has received attention in the literature. Brunnermeier (2009), Coval et al. (2009), and Ellis and Piccione (forthcoming), examine the effects of such misperceptions on financial markets. The contribution of our paper to this literature is as follows. First, our general framework rationalises how rating agencies and investment firms had coped with risky and complicated assets. Second, we show that the neglect of correlation is fundamental to rating of complex assets. Third, our notion of correlation capacity is general and encompasses specific assumptions on distributions and copulas considered in the literature to assess CDO riskiness for example.<sup>7</sup> It is particularly attractive when the modeler does not know the exact details of the environment. Fourth, we make the connection between the uncertainty that exists about these types of correlations, as documented by papers such as Duffie et al (2009), and ambiguity. In our applications we use ambiguity to examine decision making by the “pessimists” in these markets. Specifically, we show that such financial decisions, sometimes too cautious and sometimes too risky, can be characterized by a tension between cautiousness that potentially arises from ambiguity aversion and Naive-Bayes updating rules.<sup>8</sup>

---

<sup>7</sup>For a recent example of using copulas to consider the rating of a CDO see Wang et al (2009).

<sup>8</sup>There is a literature in finance looking at the role of ambiguity in investment as for example in Easley and O'hara (2009). We highlight the ambiguity over correlation in information sources. A related paper is Jiang and Tian (2016) analyze a financial market in which investors have ambiguity about the correlation between assets. Using a Gaussian distribution framework, they assume that max-min investors differ in the set of positive correlations they entertain. They derive results relating to the volume of trade and asset prices.

Our paper is related to Sobel (2014) who asks whether one can rationalise a given set of data (pre and post-deliberation actions) by finding some information structure that would rationalise this data. An agent in our model combines forecasts as the modeler in Sobel (2014). We differ by restricting of the set of possible information structures using the notion of pointwise mutual information and show that this implies meaningful constraints on the set of post-deliberation rationalisable beliefs. Finally, our approach is also related to the social learning literature (see Bikhchandani et al 1992 and Banerjee 1992). In social learning models, individuals may not extract all private information of others from their actions.<sup>9</sup> Our model highlights the possibility that even if all private information was available and shared, it would still be insufficient; knowing for example all signals and marginal distributions is not enough to understand potential correlation in the joint information structures.

We proceed as follows. The next section presents the theoretical model of combining forecasts. In Section 3 we present our main result, Theorem 1, which characterizes the set of rationalisable beliefs as a function of correlation capacity. In Section 4 we consider the implication of Theorem 1 to individual behaviour in a simple investment model in financial markets, and to CDO rating. We provide some technical extensions in Section 5 and conclude in Section 6. All proofs that are not in the text are relegated to an Appendix.

## 2 The model

In this Section we present a theoretical model in which we define what an agent observes, and how she perceives correlation. We then use this model to derive the set of predictions an agent can reach when combining forecasts, and consider different applications to see how this can affect behaviour.

We consider two types of correlation that the agent might contemplate. One correlation is across the different elements of the state. For example, when the returns of assets are correlated, or when the voting outcome across US states depends on a common shock. Another sort of correlation could be across the observed forecasts. For example, if pollsters' strategies systematically neglect parts of the population across states, this will imply a correlation across states. Similarly, banks that conduct stress tests may persistently ignore the same type of information. More generally, in many environments, both types of correlation may arise.

---

<sup>9</sup>Similarly, in the naïve social learning literature (Bohren 2014, Eyster and Rabin 2010, 2014, Gagnon-Bartsch and Rabin 2015), individuals interpret, often incorrectly, the actions of predecessors as truthful indications of their signals, and typically do not take into account that others may have learnt from others' signals.

## 2.1 Information

We first describe the information that the agent knows, which includes the state space, some prior knowledge, and observed forecasts.

The agent knows the following aspects of the environment:

1. The state space. The state is a  $n$ -tuple vector  $\boldsymbol{\omega} = \{\omega_1, \dots, \omega_n\}$ ,  $\omega_i \in \Omega$ ,  $\boldsymbol{\omega} \in \Omega^n$ , where  $n \geq 1$  and  $\Omega$  is a finite space (In Section 5.2 we consider the case of continuous distributions which may be more suitable for some applications). The agent knows the state space.
2. Priors. The agent only knows the marginal prior distributions,  $p_i(\omega_i)$ , over all elements  $i \in N$ .
3. Observed forecasts. The agent also has access to  $K$  observable forecasts. Specifically, there are  $k_i$  forecasts about each  $\omega_i$ , so that  $\sum_{i=1}^n k_i \equiv K$ . A typical forecast  $k_i^j$  is a (full support) probability distribution,  $q_i^j(\omega_i)$ , over  $\Omega$ . Let  $\mathbf{q}$  denote the vector of the  $K$  observable forecasts.

Allowing for multiple forecasts per element of the state will allow us to focus both on correlations between forecasts (even when  $n = 1$ ) as well as on correlations across the different elements of the state (when  $n > 1$ ).

The agent will combine these forecasts and his prior(s) to reach a set of predictions about the state. A prediction about the state is a probability distribution  $\eta$  over  $\Omega^n$  and we are interested in the set of rationalisable predictions that we define formally below.

## 2.2 Rational predictions

To combine forecasts into rationalisable predictions, the agent will need to consider the process according to which the observed forecasts were derived. Remember that the agent observes  $\mathbf{q}$ , the  $K$ -tuple vector of forecasts, and that he only knows the marginal prior distributions. To rationalise a prediction based on this observation, the agent considers a joint information structure.

A *joint information structure* is a vector  $(S, \Omega^n, p(\boldsymbol{\omega}), \hat{q}(\mathbf{s}, \boldsymbol{\omega}))$  consisting of:

1. A joint prior distribution,  $p(\boldsymbol{\omega})$  of which the marginal on element  $i$  is  $p_i(\omega_i)$ .
2. A set of  $K$ -tuple vectors of signals  $S = \prod_{i=1}^n \prod_{j=1}^{k_i} S_i^j$ , where  $S_i^j$  is finite and denotes the set of signals for information source  $j$  about element  $i$ .
3. A joint probability distribution of signals and states,  $q(\mathbf{s}, \boldsymbol{\omega})$ , where  $\mathbf{s} \in S$ . Specifically, let  $q(\mathbf{s}, \boldsymbol{\omega}) = p(\boldsymbol{\omega})q(\mathbf{s}|\boldsymbol{\omega})$ , where  $q(\mathbf{s}|\boldsymbol{\omega})$  is the distribution over signals generated by the state  $\boldsymbol{\omega}$ .

In a joint information structure, both the elements of the state and the signals generating the different information sources could be correlated. In particular, the prior  $p(\boldsymbol{\omega})$  might



be different than  $\prod_{i=1,\dots,n} p_i(\omega_i)$  and the function  $q(\mathbf{s}|\boldsymbol{\omega})$  may differ from  $\prod_{i=1,\dots,n} \prod_{j \in K_i} q_i^j(s_i^j|\omega_i)$ , where  $q_i^j(s_i^j|\omega_i)$  is the marginal of  $q(\mathbf{s}|\boldsymbol{\omega})$  on element  $i$  and signal  $j$ .<sup>10</sup>

We are now ready to define formally a rationalisable prediction:

**Definition 1:** *A joint information structure  $(S, \Omega^n, p(\boldsymbol{\omega}), \hat{q}(\mathbf{s}, \boldsymbol{\omega}))$  rationalizes a prediction  $\eta(\cdot)$ , given  $\mathbf{q}$ , if there exists  $\mathbf{s} = \{s_1^1, s_1^2, \dots, s_1^{k_1}, \dots, s_2^1, \dots, s_n^{k_n}\} \in S$  such that  $q_i^j(\omega_i) = \Pr(\omega_i|s_j^j) = \frac{p_i(\omega_i)\hat{q}_i^j(s_i^j|\omega_i)}{\sum_{v \in \Omega} p_i(v)\hat{q}_i^j(s_i^j|v)} \forall j \in K_i, i \in N$ , and  $\eta(\boldsymbol{\omega}) = \Pr(\boldsymbol{\omega}|\mathbf{s}) = \frac{p(\boldsymbol{\omega})\hat{q}(\mathbf{s}|\boldsymbol{\omega})}{\sum_{v \in \Omega^n} p(v)\hat{q}(\mathbf{s}|v)}$ .*

In other words, the agent perceives some  $\hat{q}(\mathbf{s}|\boldsymbol{\omega})$ , prior  $p(\boldsymbol{\omega})$ , and a particular realisation of signals  $\mathbf{s}$ , such that given the marginal of the prior per state  $p_i(\omega_i)$ , the marginal of the joint information structure  $\hat{q}_i^j(s_i^j|\omega_i)$  derived from the joint distribution  $\hat{q}(\mathbf{s}|\boldsymbol{\omega})$ , and the relevant element of  $\mathbf{s}$ ,  $s_j^j$ , all the forecasts  $q_i^j(\omega_i)$  can be rationalised by Bayes rule.

It is straightforward to show that if the set of feasible  $q(\mathbf{s}, \boldsymbol{\omega})$  is not restricted, any belief can be rationalised (see also Example 2 below). We now consider how the agent perceives correlation in order to restrict the set of feasible  $q(\mathbf{s}, \boldsymbol{\omega})$  in a meaningful way.

## 2.3 Correlation capacity

We now provide a simple and useful one-parameter characterization for the perception of correlation. In particular we formalize a notion of ‘‘correlation capacity’’, i.e., the capacity of the agent to entertain the possibility of correlation between information structures. To this end, we use the exponent of the *pointwise mutual information* (ePMI) to define possible bounds on the correlation between information structures in the minds of individuals. Specifically, we assume the following:

**Assumption A1: (Correlation capacity constraints)** *The agent is characterised by a parameter  $1 \leq a < \infty$ , and only considers joint information structures,  $(S, \Omega^n, p(\boldsymbol{\omega}), \hat{q}(\mathbf{s}, \boldsymbol{\omega}))$ , such that at any state  $\boldsymbol{\omega} \in \Omega^n$  and for any vector of signals  $\mathbf{s} \in S$ ,*

$$\frac{1}{a} \leq \frac{q(\mathbf{s}, \boldsymbol{\omega})}{\prod_{i \in N} p_i(\omega_i) \prod_{j \in K_i} q_i^j(s_i^j|\omega_i)} \leq a.^{11}$$

Thus individuals can only consider environments in which the ePMI,  $\frac{q(\mathbf{s}, \boldsymbol{\omega})}{\prod_{i \in N} p_i(\omega_i) \prod_{j \in K_i} q_i^j(s_i^j|\omega_i)}$ , is bounded, and the parameter  $a$  determines the scale of the correlation they consider. The formulation above is general enough to encompass all specific distributions and copulas considered in the literature (bar those with full correlation as  $a$  is bounded). We now explain in more detail how this relates to actual examples of correlated information structures.

<sup>10</sup>In the main part of the analysis we assume that the agent only observes the forecasts  $\mathbf{q}$ . In Section 5.1 we show that our results are robust to the agent also observing the signals and the marginal information structures of the different sources. Also in this case there would still be different types of correlation across the marginal information structures.

<sup>11</sup>All the results can be easily generalized if instead of the lower bound  $\frac{1}{a}$  we use some finite  $b < 1$ .

**Pointwise mutual information: theoretical background.** PMI was suggested by Church and Hanks (1991) and is used in information theory and text categorization or coding, to understand how much information one word or symbol provides about the other, or to measure the co-occurrence of words or symbols. Let  $f(x_1, \dots, x_n)$  be a joint probability distribution of random variables  $\tilde{x}_1, \dots, \tilde{x}_n$ , with marginal distributions  $f_i(\cdot)$ . The pointwise mutual information (PMI) at  $(x_1, \dots, x_n)$  is  $\ln[\frac{f(x_1, \dots, x_n)}{\prod_i f_i(\cdot)}]$ . For example, for two variables, it can also be written as

$$\ln[\frac{f(x_1, x_2)}{\prod_i f_i(\cdot)}] = h(x_1) - h(x_1|x_2)$$

where  $h(x_1) = -\log_2 \Pr(X_1 = x_1)$  is the self information (entropy) of  $x_1$  and  $h(x_1|x_2)$  is the conditional information.

Summing over the PMIs, we can derive the well known measure of mutual information,  $MI(X_1, X_2) = \sum_{x_1 \in X_1} \sum_{x_2 \in X_2} f(x_1, x_2) \ln[\frac{f(x_1, x_2)}{\prod_i f_i(\cdot)}] = H(X_1) - H(X_1|X_2)$ , which can be shown to be always non-negative as it equals the amount of uncertainty about  $X_1$  which is removed by knowing  $X_2$ . We can also express mutual information by using the definition of Kullback-Leibler divergence between the joint distribution and the product of the marginals:

$$MI(X_1, X_2) = D_{KL}(f(x_1, x_2)|f_1(x_1)f_2(x_2)),$$

and it can therefore capture how far from independence individuals believe their information structures are. For our purposes, the local concept of the PMI is a more suitable concept than the MI, as we are looking at *ex-post* rationalizations given some set of signals.<sup>12</sup>

The concept of the PMI is closely related to standard measures of correlation and specifically it implies a bound on the *concordance* between information structures. In the Appendix we show (Proposition B1) how a bounded PMI translates to a bounded concordance measure.<sup>13</sup> In our analysis we use the ePMI, which is the exponent of the PMI, i.e.,  $\frac{f(x_1, \dots, x_n)}{\prod_i f_i(\cdot)}$ .

**Examples:** We use the examples below to illustrate the relation between ePMI, the level of  $a$ , and correlation.

**Example 1:** Assume  $\Omega = \{0, 1\}$ , two states,  $\omega_1$  and  $\omega_2$ , and one information source per state, so that  $n = 2$  and  $k_i = 1$ , with  $K = 2$ . Assume that the agent thinks that the joint signal structure  $q(s|\omega)$  satisfies independence, but that the prior satisfies correlation, as described in the following symmetric matrix, where  $p(0) + p(1) = 1$ :

$$\begin{array}{cc} & \omega_2 = 0 & \omega_2 = 1 \\ \omega_1 = 0 & p(0)^2 + \varepsilon & p(0)p(1) - \varepsilon \\ \omega_1 = 1 & p(0)p(1) - \varepsilon & p(1)^2 + \varepsilon \end{array}$$

<sup>12</sup>The PMI therefore does not distinguish between rare or frequent events.

<sup>13</sup>The most common measure of concordance is Spearman's rank correlation coefficient or Spearman's  $\rho$ , a nonparametric measure of statistical dependence between two variables. A perfect Spearman correlation of +1 or -1 occurs when each of the variables is a perfect monotonic function of the other. In the Appendix we show that there is a  $0 < \bar{\rho} < 1$  such that any joint information structure with bounded PMI's has a Spearman's  $\rho$  in  $[-\bar{\rho}, \bar{\rho}]$ . This also implies that we can put bounds directly on the copula.

When  $\varepsilon = 0$ , we have ePMI equals 1 at any state. If  $\varepsilon$  is positive (and small enough), then we have positive correlation across the states, whereas if  $\varepsilon$  is negative, we have negative correlation. Consider a positive  $\varepsilon$ . Note now that the ePMI at  $\omega = \{0, 0\}$  is  $\frac{p(0)^{2+\varepsilon}}{p(0)^2} > 1$ , whereas the ePMI at  $\omega = \{0, 1\}$  is  $\frac{p(0)p(1)-\varepsilon}{p(0)p(1)} < 1$ .

**Example 2:** Assume  $\Omega = \{0, 1\}$ , only one element for the state,  $\omega_1$ , and two information sources. Thus  $n = 1$ ,  $k_1 = 2$  and  $K = 2$ . Assume that the agent perceives the following symmetric information structure with two signals:

$$\begin{array}{cccccc} \omega_1 = 0 & s_2 = s^* & s_2 = s^{**} & \omega_1 = 1 & s_2 = s^* & s_2 = s^{**} \\ s_1 = s^* & q(s^*|0) - q_0 & q_0 & , & s_1 = s^* & q(s^*|1) - q_1 & q_1 \\ s_1 = s^{**} & q_0 & q(s^{**}|0) - q_0 & & s_1 = s^{**} & q_1 & q(s^{**}|1) - q_1 \end{array}$$

Note that when  $a$  is not bounded, we can construct information structures which rationalize strong individual forecasts that the state is 1, but, a certain combined forecast that the state is 0. For example, if  $q_1 = q(s^*|1)$ , and  $q_0 \rightarrow 0$ , then upon receiving  $s^*$ , each information source believes that the state is very likely to be 1. However, upon observing both forecasts, the agent is convinced according to the information structure above and these assumptions that the probability that the state is 1 is zero. This follows as the ePMI at  $(s^*, s^*|\omega_1 = 1)$  is 0, which demands an unbounded  $a$ . However, when  $a$  is bounded, this cannot arise. As in Example 1, for example when  $q_1 = q(s^*|1)q(s^{**}|1) - \varepsilon$ , we have positive correlation in signals in state 1. And the higher is  $a$ , the higher is the level of positive correlation that we can attain.

**Example 3\*:** Our model is finite for simplicity but can be easily extended to a continuous space. We analyze an example of this in Section 5.2, which we now introduce. Specifically, consider  $\Omega = \mathbb{R}$ , one state  $\omega$ , two information sources, and a known prior  $p(\omega)$ . Suppose that, among others, the agent perceives symmetric marginal distributions  $g(s^1|\omega)$  and  $g(s^2|\omega)$ , with PDFs  $G(s^1|\omega)$  and  $G(s^2|\omega)$  respectively, and the following family of joint information structures, constructed according to the Morgenstern transformation:

$$g(s^1, s^2|\omega) = [1 + \alpha_\omega(2G(s^1|\omega) - 1)(2G(s^2|\omega) - 1)]g(s^1|\omega)g(s^2|\omega).$$

In this family,  $\alpha_\omega > (<)0$  signifies positive (negative) correlation. For this family to hold, it has to be that  $|\alpha_\omega| \leq 1$ . Furthermore, to satisfy the ePMI constraints for some  $a$ , we also need that at any  $s^1, s^2, \omega$ :

$$\frac{1}{a} \leq [1 + \alpha_\omega(2G(s^1|\omega) - 1)(2G(s^2|\omega) - 1)] \leq a$$

Thus, a higher  $a$  will allow for more positive correlation across information sources if  $\alpha_\omega > 0$ , and for more negative correlation across information sources if  $\alpha_\omega < 0$ .

**Complexity and correlation capacity:** More generally, as the examples above illustrate, whenever the ePMI at some point is greater than 1, it has to be smaller than 1 at

another set of states or set of signals for the same state. Thus whenever one perceives the ePMI to be less than 1 he has to perceive it to be greater than 1 as well, which implies that perceiving the ePMI at 1 is in some sense the simplest possibility and hence is always in the set. It is also easy to see from the above that the higher is  $a$ , the more priors and joint information structures can be considered as we can increase  $\varepsilon$  and still maintain as in Example 1,  $\frac{p(0)^{2+\varepsilon}}{p(0)^2} < a$ ,  $\frac{p(0)p(1)-\varepsilon}{p(0)p(1)} > \frac{1}{a}$ , and so on.

The methodology behind assumption A1 is similar to the rational inattention model in Sims (2003) which allows individuals to consider models of the world with finite Shannon capacity. We think about  $a$  as a parameter describing the cognitive capacity of the agent, although in different environments individuals may be able to have different such capacities (for example,  $a$  may depend on the number of sources  $n$ ). Globally, it is more difficult to perceive a complicated correlation structure compared to independence. Even locally, increasing  $a$  implies that individuals can imagine models of information structures in which not only the signals of all sources become slightly more correlated, but cases in which the signals of some become less correlated and signals of others become much more correlated. Such asymmetries can be regarded as more complex structures.

Note that A1 is general enough so that when pitted against the fully rational individual who knows the joint information structure, individuals with information capacity  $a$  may potentially neglect correlation or perceive too much correlation. We will discuss this in specific applications later on.

**Rationalisable beliefs with correlation capacity:** Let the set  $C(a, \mathbf{q})$  be the set of beliefs  $\eta(\cdot)$  that are rationalisable, as in Definition 1, given the vector of forecasts  $\mathbf{q}$ , by information structures that satisfy Assumption 1 for a correlation capacity  $a$ .

Our main result below characterizes  $C(a, \mathbf{q})$ . We will show that focusing on correlation capacity provides a useful and simple characterization of beliefs. We will also show that in some environments, this set of beliefs can become very small even when individuals consider large degrees of correlation.

### 3 Combining forecasts with correlation capacity

In this Section we characterise  $C(a, \mathbf{q})$ , the set of post-communication rationalisable beliefs of the agent that are derived from information structures that satisfy A1 for  $a$ .

Suppose first that there is only one state and one source of information about the state. In other words, when  $n = 1$  and  $K = 1$ . It is then easy to see that:

**Lemma 0:** *When  $n = 1$  and  $K = 1$ , for any  $a$ , the only rationalisable belief on the state is the observed prediction  $q(\omega)$ .*

For any information source that the agent can imagine, we know that  $\frac{p(\omega)q(s|\omega)}{\sum_v p(v)q(s|v)}$  has to

equal  $q(\omega)$  by rationalisability. Correlation does not play a role here as there is only one element in the state and one information source.

Let us consider now the case in which there is still one state, so that  $n = 1$ , but several information sources. Suppose that the individual has no correlation capacity, so that  $a = 1$ . As  $n = 1$ , the agent knows the prior on the state. As  $a = 1$ , the agent can only perceive independent joint information structures for the signals. Define an agent who believes that information sources are independent as a Naive-Bayes (NB) agent. The agent then needs to “imagine” such information structures and signals that will deliver the predictions observed. As with one source, the predictions are sufficient statistics for the signals of the information sources. We then have:

**Lemma 1a:** *When  $n = 1$  and  $K > 1$ , the set of rationalisable beliefs of a NB agent is a singleton and the unique belief is  $\eta^{NB}(\omega)$  satisfying:*

$$\frac{\eta^{NB}(\omega)}{\eta^{NB}(\omega')} = \frac{\prod_{j=1}^K q^j(\omega)}{\frac{p(\omega)^{K-1}}{p(\omega')^{K-1}}}$$

By Bayes rule, the predictions are sufficient statistics, and the agent only needs to take into consideration that each source had used the prior in its update. To make sure that the prior is not taken into account too much the multiplication of the forecasts is then divided by the prior.

It is straightforward to extend the above to many states. Continue with the benchmark of the NB agent, we then have:

**Lemma 1b:** *When  $n > 1$ , the set of rationalisable beliefs of a NB agent ( $a = 1$ ) is a singleton and the unique belief is  $\eta^{NB}(\omega)$  satisfying, for any  $\omega = (\omega_1, \dots, \omega_i, \dots, \omega_n)$  and  $\omega' = (\omega'_1, \dots, \omega'_i, \dots, \omega'_n)$ :*

$$\frac{\eta^{NB}(\omega)}{\eta^{NB}(\omega')} = \frac{\prod_{i=1}^n \frac{\prod_{j=1}^{k_i} q_i^j(\omega_i)}{p(\omega_i)^{k_i-1}}}{\prod_{i=1}^n \frac{\prod_{j=1}^{k_i} q_i^j(\omega'_i)}{p(\omega'_i)^{k_i-1}}}$$

As the agent only perceives independent priors and information sources within and across the elements of the state, the result of Lemma 1a extends. To see this formally, note that by Bayes rule, any belief needs to satisfy the following likelihood ratio for some vector of signals and marginal information structures:

$$\frac{\prod_{i=1}^n (p_i(\omega_i) \prod_{j=1}^{k_i} q_i^j(s_i^j | \omega_i))}{\prod_{i=1}^n (p_i(\omega'_i) \prod_{j=1}^{k_i} q_i^j(s_i^j | \omega'_i))}$$

By rationalisability however:

$$\frac{\prod_{i=1}^n (p_i(\omega_i) \prod_{j=1}^{k_i} q_i^j(s_i^j|\omega_i))}{\prod_{i=1}^n (p_i(\omega'_i) \prod_{j=1}^{k_i} q_i^j(s_i^j|\omega'_i))} = \frac{\prod_{i=1}^n (p_i(\omega_i) \prod_{j=1}^{k_i} \frac{q_i^j(\omega_i) \sum_{v_i} p_i(v_i) q_i^j(s_i^j|v_i)}{p_i(\omega_i)})}{\prod_{i=1}^n (p_i(\omega'_i) \prod_{j=1}^{k_i} \frac{q_i^j(\omega'_i) \sum_{v_i} p_i(v_i) q_i^j(s_i^j|v_i)}{p_i(\omega'_i)})} = \frac{\prod_{i=1}^n \prod_{j=1}^{k_i} q_i^j(\omega_i)}{\prod_{i=1}^n \frac{p_i(\omega_i)^{k_i-1}}{p_i(\omega'_i)^{k_i-1}}}$$

What if the agent has capacity for correlation? We now proceed to provide the main result.

### 3.1 The main result

**Theorem 1:** *With  $n$  states and  $K$  information sources,  $\eta(\cdot) \in C(a, \mathbf{q})$  is rationalisable if and only if it satisfies*

$$\frac{\eta(\boldsymbol{\omega})}{\eta(\boldsymbol{\omega}')} = \frac{\lambda_{\boldsymbol{\omega}} \eta^{NB}(\boldsymbol{\omega})}{\lambda_{\boldsymbol{\omega}'} \eta^{NB}(\boldsymbol{\omega}')},$$

for a vector  $\boldsymbol{\lambda} = (\lambda_{\boldsymbol{\omega}})_{\boldsymbol{\omega} \in \Omega^n}$  satisfying  $\lambda_{\boldsymbol{\omega}} \in [\frac{1}{a}, a]$  for all  $\boldsymbol{\omega}$ . Thus the minimum (maximum) belief on state  $\boldsymbol{\omega}$  is derived when  $\lambda_{\boldsymbol{\omega}} = \frac{1}{a}$  ( $a$ ) and for all other  $\mathbf{v}$ ,  $\lambda_{\mathbf{v}} = a$  ( $\frac{1}{a}$ ). (iii) The set  $C(a, \mathbf{q})$  is compact and convex.

When  $a > 1$ , the set of beliefs is not unique. When the agent is able to consider more complex environments, she can rationalise a larger set of beliefs about the state of the world.

As the NB agent already accounts for rationalisability, when considering an agent with  $a > 1$ , the key addition is the larger set of ePMI constraints. It is therefore easy to see the necessary part of the proof. Specifically, note that for any  $\mathbf{s}$  and a joint information structure, the ePMI constraints imply:

$$\frac{\frac{1}{a} \prod_{i=1}^n (p_i(\omega_i) \prod_{j=1}^{k_i} q_i^j(s_i^j|\omega_i))}{a \prod_{i=1}^n (p_i(\omega'_i) \prod_{j=1}^{k_i} q_i^j(s_i^j|\omega'_i))} \leq \frac{p(\boldsymbol{\omega})q(\mathbf{s}|\boldsymbol{\omega})}{p(\boldsymbol{\omega}')q(\mathbf{s}|\boldsymbol{\omega}')} \leq \frac{a \prod_{i=1}^n (p_i(\omega_i) \prod_{j=1}^{k_i} q_i^j(s_i^j|\omega_i))}{\frac{1}{a} \prod_{i=1}^n (p_i(\omega'_i) \prod_{j=1}^{k_i} q_i^j(s_i^j|\omega'_i))}$$

which by rationalisability, as above, implies:

$$\frac{\frac{1}{a} \eta^{NB}(\boldsymbol{\omega})}{a \eta^{NB}(\boldsymbol{\omega}')} \leq \frac{p(\boldsymbol{\omega})q(\mathbf{s}|\boldsymbol{\omega})}{p(\boldsymbol{\omega}')q(\mathbf{s}|\boldsymbol{\omega}')} \leq \frac{a \eta^{NB}(\boldsymbol{\omega})}{\frac{1}{a} \eta^{NB}(\boldsymbol{\omega}')}.$$

The proof in the Appendix shows the sufficiency of the characterisation and that the set of beliefs is convex. We show sufficiency by constructing an information structure that yields each beliefs in the set and satisfies the constraints. Convexity is not trivial in our environment; due to the nature of the ePMI constraints, one cannot simply take a convex combination of joint information structures to rationalize a convex combination of beliefs. To prove convexity we therefore need to use our characterization to find a new information structure to rationalize any convex combination of beliefs in  $C(a, \mathbf{q})$ .

### 3.2 The Naive-Bayes and cautiousness effects

The Theorem shows that there are two sufficient statistics that allow us to characterize the set of beliefs held by the agent: her correlation capacity  $a$  and the NB posterior  $\eta^{NB}(\cdot)$ . Thus, while the agent is faced with a complicated environment, her Bayesian combined forecasts when she considers different information structures can be derived with a simple heuristic-like behavior. The agent needs to consider the Naive-Bayes benchmark, as if she neglects correlation, and adjust this by different “scenarios” as determined by  $a$ . This is also helpful for the modeler as we had not made any specific assumptions on distributions.

This simple characterisation allows us to make the following observations:

**Observation 1:** *Suppose that  $\eta^{NB}(\cdot)$  is close to being degenerate so that  $\eta^{NB}(\omega')$  is close to 1 for some  $\omega'$ . Then the agent has beliefs that are close to 1 on  $\omega'$ .*

While agents may have different correlation capacities, the Naive-Bayes beliefs,  $\eta^{NB}(\cdot)$ , may overwhelm any correlation capacity. As beliefs are centered around the Naive-Bayes benchmark, if  $\eta^{NB}(\omega)$  is close to being degenerate on some  $\omega'$  and  $a$  is bounded, then  $C(a, q)$  collapses to  $\eta^{NB}(\omega)$  for all  $\omega$ . In other words, if the independent benchmark is perceived as informative, it overcomes any (finite) correlation capacity. Correlation neglect can arise endogenously then. While it is important to have a bounded  $a$  in the result above, it can generalise to environments in which  $a$  converges to infinity slower than the NB benchmark converges to be degenerate.

One case in the NB benchmark is likely to become degenerate is when  $K$  becomes large, which accords with either many elements in the state or with a large number of information sources. Note that the Naive-Bayes benchmark -while relying on many pieces of information- can still differ substantially from the rational belief given the true information structure.

As agents’ beliefs will collapse to a singleton, agents will also be more confident in their combined predictions as their set of rationalisable beliefs will shrink. This accords with the experiments of Budescu and Yu (2007) who show that decision makers were more confident when the advisors’ estimates were more extreme and advisors were in higher agreement with each other.

**Observation 2:** *If  $a^i < a^j$ , then  $C(a^i, q) \subset C(a^j, q)$ .*

This observation focuses on the second sufficient statistic, which is the level of correlation capacity. Greater correlation capacity or allowing for greater complexity results in a larger set of feasible beliefs. This is an interesting relation we unravel between correlation and confidence; capacity for correlation or to consider complex environments will reduce confidence in the sense that individuals may not be sure what is the right belief. One way to interpret this is that individuals with larger  $a$  will have greater ambiguity over the state of the world.<sup>14</sup> This accords with the experiments of Maines (1990) who shows that subjects

---

<sup>14</sup>Thinking about fundamentals, one can start from a model with ambiguity over the joint information

exhibited under-confidence when they combined forecasts, because they assumed high redundancy among forecasts. In our model, individuals with low  $a^i$  will neglect correlation and will be relatively confident in the sense of having a small set of beliefs, and individuals with high  $a$  will face larger uncertainty.

For any  $\eta^{NB}(\cdot)$ , a high enough  $a$  can generate a low enough minimum belief in this set. Along with ambiguity aversion, or alternatively with pessimists taking hold in organizations, this can result in a more cautious behaviour. Thus the level of  $a$  will create the *cautiousness* effect. This effect, as we illustrate in the next Section, can explain pessimistic behavior in financial markets when investors believe they face unknown levels of correlation.

## 4 Applications

The applications below highlight the interplay between the two effects described above. We consider two application, each highlighting the different uncertainties over correlation. In the first application we consider, we focus on a simple state with one element only,  $n = 1$ , but allow for many information sources and correlation across these sources. We highlight a financial example in which individuals can choose whether to invest in a risky or safe asset, and their observations of others' investments can be interpreted as others' predictions. We use our analysis to show how the cautiousness and the NB effects combine to create booms and busts in investment behaviour. Our second application focuses on the case in which there is correlation across the elements of the state (so that  $n > 1$ ). Specifically we consider how a CDO is rated when it is composed of many loans. In this application we show how complex CDOs would be -sometimes wrongly- highly rated even when analysts take correlation into consideration.

### 4.1 Application I: Booms and busts

The 2008 financial crisis has shown that markets can move quickly from booms to busts. Specifically, from years of large investment volumes and speculations, to years of little investment and skepticism. One explanation for the scale of the 2008 crisis was that investors did not appreciate the level of correlation across assets. The large degree of skepticism and cautiousness following the crisis can however be explained by re-calibrating and perhaps even over-shooting of the levels of possible correlation, by investors realizing that they have little knowledge of the true model of correlation across assets.

In this Section we consider a simple investment model which we apply to booms and busts in financial markets. We focus attention on the case of  $n = 1$  with a binary realisation, that is, whether the economy is good for investments or not- and focus on the correlation across

---

structures. The larger is this ambiguity (stemming from a larger  $a$ ), the larger is the resulting ambiguity over the state of the world upon observation of forecasts. We consider this model in Appendix B.



information sources. That is, some investors may realise that the information they receive may be correlated.<sup>15</sup>

In this application we generalise the theoretical model presented in Section 2 in two ways. First, we assume, as in financial herding models, that each agent does not only combine forecasts but also receives a signal of her own. Second, we assume that agents have ambiguity when facing a set of beliefs and discuss the effects of ambiguity aversion.<sup>16</sup>

**A simple investment model:** Assume a binary model with two equally likely states of the world,  $\omega \in \{0, 1\}$ . Assume that there is a safe asset which provides the same returns  $L > 0$  at any state, and a risky asset which provides 0 at state 0 and  $H > L$  in state 1. The agent has one unit of income to invest which she can split across these two assets. Assume a standard concave utility  $V(\cdot)$  of wealth. Thus in this simple model the agent would invest a higher share in the risky asset the higher are her beliefs that the state is 1.<sup>17</sup>

There are  $K$  informed investors, each receiving a signal  $s^j$  on  $\omega$ , knowing the marginal  $q^j(s^j|\omega)$ , and updating their prediction to  $q^j(\omega)$ . These  $K$  investors act in the first period and invest according to  $q^j(1)$ .

After observing the investments in the first period, a vector  $\mathbf{q}$ , the investors update their beliefs and can adjust their investments in the second period. Specifically, as in the first period the share of units that is allocated to the risky asset/s would be higher for any investor the higher is the probability she believes that the state is 1. Therefore, if we assume that the fraction of the investment in the risky asset is observable, then individuals can backtrack the beliefs of others,  $q^j(1)$ .<sup>18</sup>

We show in Appendix B how all the analysis and the results reported above remain also in the case in which the agent who combines forecasts has information of her own, as we assume here.<sup>19</sup> We can therefore use the results above to assess how investment will change

---

<sup>15</sup>Note that the literature on herding in financial markets typically focuses on the idea that actions do not perfectly reveal information or beliefs. In our model we assume that the actions of investors reveals their beliefs perfectly. However, individuals are puzzled with regard to how to interpret these beliefs given that there may be correlation in the investors' sources of information.

<sup>16</sup>There is recently an increasing literature on ambiguity in finance. See for example Easley and O'hara (2009) and Jiang and Tian (2016).

<sup>17</sup>Here we abstract from prices. The analysis could be extended to consider prices and market makers (see the discussion below).

<sup>18</sup>Of course this assumption is somewhat extreme but it not necessary and is made here for simplicity. One can easily assume a weaker version in which just the quantity invested is observed. In that case, after observing investments, agents will not infer the beliefs exactly but rather will be able to compute lower bounds on these beliefs.

<sup>19</sup>Specifically, we need to show that when the individual receives her signal, her uncertainty about the joint information structure (but her knowledge of her marginal distribution) leads her still to a unique belief, which is straightforward to show. Another issue is that as she needs her marginal to update her belief to  $q^j(\omega)$ , the set of rationalisable beliefs may depend on her marginal. One possibility is to assume that when combining forecasts the investors only remember their posterior belief and not the process that lead to it. Alternatively we can conduct the same analysis as in Theorem 1, with the knowledge of the marginals and

in the second period.

Note that individuals can differ in our model in their correlation capacity but also in how they behave when faced with sets of beliefs. Some may have a distribution over the different correlation structures, some may face ambiguity, some may be optimists etc. To focus on correlation capacity, we fix here the behavioural assumptions. Specifically, we assume that in the face of a set of beliefs agents have ambiguity aversion and concretely, that when individuals are faced with ambiguity, they use the max-min preferences as in Gilboa and Schmeidler (1989). Note that given the convexity of the set of beliefs  $C(a, \mathbf{q})$ , we can use other attitudes towards ambiguity to generate similar results. Moreover, when correlation neglect arises endogenously it arises for any behavioural assumption as the set of beliefs shrinks to a singleton.

Following exposure to multiple forecasts and ambiguity aversion, an individual  $j$  will base her investment decision in the second period on the belief which minimises her utility:  $\min_{\eta^j(1) \in C(a^j, \mathbf{q})} \eta^j(1)$ .

Given Theorem 1, and as we are in the binary model with only two states of the world, we can further simplify  $\eta^{NB}(\cdot)$ . Let  $\hat{q}(1)$  be the belief such that  $\frac{\hat{q}(1)}{1-\hat{q}(1)}$  is the geometric average of  $\{\frac{q^j(1)}{1-q^j(1)}\}_{j \in K}$ , i.e.,  $\frac{\hat{q}(1)}{1-\hat{q}(1)} = (\prod_{j \in K} \frac{q^j(1)}{1-q^j(1)})^{\frac{1}{k}}$ . We can now express  $\eta^{NB}(1)$  as (recall that  $k$  is the number of investors),

$$\eta^{NB}(1) = \frac{\hat{q}(1)^k}{(1-\hat{q}(1))^k + \hat{q}(1)^k}.$$

Given Theorem 1,  $j$ 's minimum combined forecast can be then written as:

$$\min_{\eta(1) \in C(a^j, \mathbf{q}(1))} \eta(1) = \frac{\frac{1}{a^j} \hat{q}(1)^k}{a^j (1-\hat{q}(1))^k + \frac{1}{a^j} \hat{q}(1)^k}. \quad (1)$$

Thus the agent invests more in the second period if and only if:

$$\frac{q^j(1)}{1-q^j(1)} < \frac{1}{(a)^2} \left( \frac{\hat{q}(1)}{1-\hat{q}(1)} \right)^k. \quad (2)$$

**Lemma 2:** *In the second period, following exposure to the same predictions  $\mathbf{q}$ : (i) an investor  $j$  who has a lower correlation capacity than investor  $j'$  will invest more in the risky asset compared with  $j'$ . (ii) For any  $k$ , there is  $\gamma > 1$  such that if  $a^j > \gamma$  then individual  $j$  will lower his investment in the risky asset. (iii) For a large enough  $k$ , if  $\hat{q}(1) > \frac{1}{2}$ , then all investors will substantially increase their investment in the risky asset and if  $\hat{q}(1) < \frac{1}{2}$ , then all will substantially decrease their investment in the risky asset.*

Part (i) follows from (2) and illustrates that different capacity for correlation can have different implications for behaviour. Specifically, individuals with less correlation capacity will behave in a more risky manner. Note that “standard” results in the literature on signals (see Section 5.1).

correlation neglect are typically of the form that individuals with more correlation neglect will take more extreme decisions, but depending on the state of the world these could be either on the risky or on the cautious side. The result above is different; it applies to any state of the world and any set of signals, and arises from the reduced ambiguity that comes with lower perception of correlation.<sup>20</sup> This result also implies that we can identify capacities for correlation from choice data, as long as there is data on behavior of individuals in the face of ambiguity. Once individuals have the same attitudes towards ambiguity, we can differentiate those with lower correlation capacity by their more risky behavior.

To see how parts (ii) and (iii) arise, recall that we have identified the cautiousness effect and the Naive-Bayes effect. If beliefs are in general pessimistic (that is, against investment), then both go in the same direction inducing a cautious behaviour following exposure to multiple sources. If on the other hand beliefs are optimistic (namely,  $\hat{q}(1) > \frac{1}{2}$ ) the two effects go in opposite direction. Whichever dominates depends on the level of correlation capacity as well as on how close is the  $\eta^{NB}(\cdot)$  to being degenerate.

When the number of forecasts is small, the Naive-Bayes effect is weak, as  $\eta^{NB}(\cdot)$  is not likely to be informative. In other words,  $(\frac{\hat{q}(1)}{1-\hat{q}(1)})^k$ , while greater than 1, is not too large. We can then always find a high enough correlation capacity so that the cautiousness effect will dominate.<sup>21</sup>

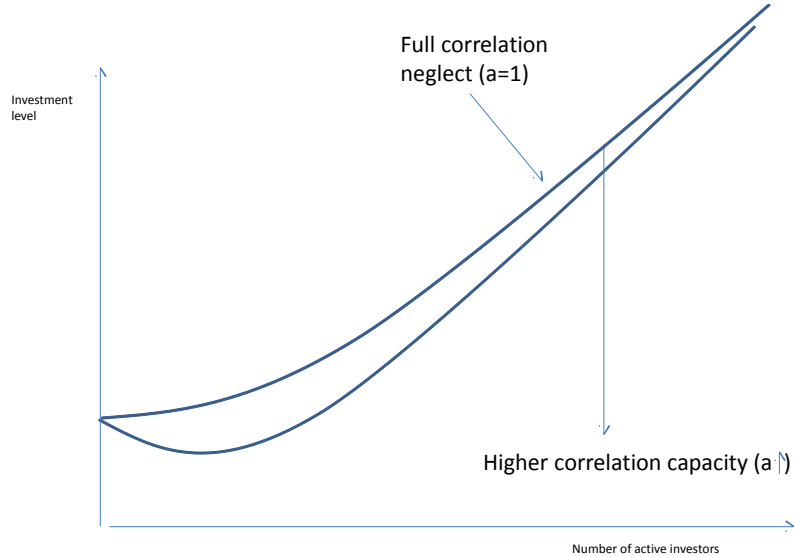
On the other hand, when the number of forecasts is large, the Naive-Bayes effect dominates. If beliefs are optimistic,  $(\frac{\hat{q}(1)}{1-\hat{q}(1)})^k$  becomes very large and this effect overcomes the cautiousness one to induce a substantial risky behaviour.

**Booms and busts:** Given Lemma 2, the following figure illustrates the behaviour of investors in the second period, for different correlation capacities (manifested in the different curves), as a function of the number of investors. The figure is constructed for all individuals starting with the same optimistic beliefs  $q^j(1) = q = \hat{q}(1) > \frac{1}{2}$ , but differing in their correlation capacities. It illustrates how some individuals with large  $a^j$  respond with a more cautious behaviour when the number of other investors is small, but with a very risky behaviour when the number of other investors grows large.

---

<sup>20</sup>Note that in a model without ambiguity, when one compares an individual with correlation neglect to one without correlation neglect, the former will have more extreme beliefs but sometimes in the risky and sometimes in the cautious direction. For example, in the papers of Glaeser and Sunstein (2009) or Ortoleva and Snowberg (2015), individuals believe that a set of signals is independently drawn rather than repeated. This implies that they have more extreme beliefs -be it high or low beliefs- compared to an individual who realizes that these signals are repeated.

<sup>21</sup>This result also has a flavour of dynamic inconsistency results in the Ambiguity literature. See Hanani and Klibanoff (2007) for updating that restricts the set of priors and avoids dynamic inconsistency, and the discussions in Al-Najjar and Weinstein (2009) and Siniscalchi (2011).



Note that when the number of investors is small, investors can diverge depending on their correlation capacity. While those with high correlation capacity will decrease their investment, due to the cautiousness effect, those with low correlation capacity will increase their investment, as they neglect correlation. The former have little confidence in their beliefs, as they face large ambiguity, while the latter are more confident.<sup>22</sup>

Given the above, we can summarise how booms and busts arise in the following corollary:

**Corollary 1:** *A boom happens in the second period when  $k$  is large and  $\hat{q}(1) > \frac{1}{2}$ . A bust happens when either  $k$  is small and  $a$  is high for most investors, or when  $k$  is large and  $\hat{q}(1) < \frac{1}{2}$ .*

One parameter that had changed after the financial crisis is investors' perception of correlation, as modeled by  $a$  in our model. Shocks such as the financial crisis that affect investor anxiety about the complexity or correlation between sources of data will raise the bar for markets to exhibit growth in investment. This fear of unknown correlation might have played a role in the post-2008 years. For any observed behavior of others, a higher value of  $a$  is more likely to decrease investment. While before the crisis investors and rating agencies had tended to ignore correlation, with optimists (low  $a$ ) winning over pessimists (high  $a$ ) when different scenarios were considered, after the crisis the worst case scenarios did not only receive more weight in the overall assessment, but were also downgraded themselves to capture a more pessimistic outlook. Thus the cautiousness effect that we had identified, as exemplified by  $a$ , had more bite in the post-crisis years.

Another element that changes in the market is the informativeness of  $\eta^{NB}(\cdot)$  which depends on the number of investors involved. A market with many investors (even small ones) is such

<sup>22</sup>This provides another explanation for group divergence as studied in Lord et al (1979): Even though individuals observe the same information they can diverge in their beliefs or actions. According to our model, these individuals will also diverge in their confidence level.

that individuals can observe many forecasts. Even if each investment is slightly optimistic, it can be aggregated to a precise and very optimistic  $\eta^{NB}(\cdot)$ , which will overshadow the cautiousness effect, as in part (iii) of Lemma 1 for the case of  $\hat{q}(1) > 1/2$ . This may respond to the investment levels pre-2008. On the other hand, once some skepticism arises, the market will consist of less investors. In this case, individuals cannot learn much, as they have too few forecasts to combine. The benchmark  $\eta^{NB}(\cdot)$  would be imprecise, and the cautiousness effect will dominate, as in part (ii) of Lemma 1. A less active market or such that consists of perhaps the same volume of trade but with lower number of investors will allow the cautiousness effect to dominate and for skepticism to escalate.

Note that most of the finance literature on booms and busts, or herding and information cascades, had focused on the case of independently distributed private information. The key idea in this literature is that actions are not sufficient statistics to beliefs. However, our framework points at a deeper problem; even if actions fully reveal beliefs,<sup>23</sup> investors will not be able to reconstruct the joint information structure. Any observation of a vector of investment will still be open to different interpretation, with some potentially interpreting it pessimistically and some optimistically. Still, we have environments in which public information will overwhelm any individual interpretation.

Finally, there are several ways in which one can extend the model above. First, we can consider a model with more than two periods; our qualitative results will still be maintained. Second, we can endogenize  $k$ , the number of active investors, which is a matter for future research.<sup>24</sup> Finally, we can extend the above analysis to include prices determined market makers, with some added assumptions which guarantee that there is asymmetric information between informed investors and market makers, as in Avery and Zemsky (1998).<sup>25</sup>

## 4.2 Application II: Rating a CDO

In this Section we provide a simple model of risk management for CDOs. Consider the case in which a CDO consists of  $n$  loans, each with a binary state of default (D) or no default (ND),  $\Omega = \{D, ND\}$ . Suppose that a particular tranche of the CDO defaults if at least a share  $\alpha$  of the individual loans included in it will default, meaning for at least  $\lceil \alpha n \rceil$  elements of the state, we have  $\omega_i = D$ .

In this application the uncertainty over correlation will be about the correlation between

---

<sup>23</sup>Or, more generally even if investors will fully reveal their signal and marginal information structures.

<sup>24</sup>For example, assume that participation costs deter individuals in the first period (disregarding their beliefs). If beliefs exhibited in the first period are relatively optimistic, and those that did not participate have low correlation capacity, they will enter in the second period (if this becomes feasible). This can increase the number of investors in the market to a high enough  $k'$  which at some point can become a sufficiently optimistic indicator to induce even those with high correlation capacity to enter. Alternatively, if the distribution of investors has high correlation capacity it may be that a low number of investors in the first period will remain so.

<sup>25</sup>See the surveys of Vayanos and Wang (2013) and Bikchandani and Sharma (2000).

the defaults of the individual loans. Therefore, we assume that there are no observable forecasts. The prior marginal probability of each loan defaulting (or  $\omega_i = D$ ) is  $p_i(\omega_i = D) = p_i$ , and therefore we have  $n$  Bernoulli trials, each with a marginal probability of  $D$  of  $p_i$ . When the trials are independent, this is a Poisson Binomial distribution. Below, when we take  $n$  to be large we will assume that  $\lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n p_i}{n} = \mu < \infty$ .

This is the simplest static model that can describe a CDO (alternatively, one can consider a dynamic probability of default, meaning a Poisson distribution, which our model can easily be extended to). Moreover, other models typically assume a particular parametric family of copulas to assess the cumulative risk of assets.<sup>26</sup> We instead describe correlation capacity without resorting to any functional forms.

By Theorem 1, for any state  $\omega$ , where  $\omega_i \in \{D, ND\}$ , we have that a belief  $\eta(\cdot)$  is in the set  $C(a, \mathbf{q})$  iff:

$$\frac{\eta(\omega)}{\eta(\omega')} = \frac{\lambda_\omega \eta^{NB}(\omega)}{\lambda_{\omega'} \eta^{NB}(\omega')} = \frac{\lambda_\omega \prod_{i=1}^n p_i(\omega_i)}{\lambda_{\omega'} \prod_{i=1}^n p_i(\omega'_i)}$$

for any  $\lambda_\omega, \lambda_{\omega'} \in [\frac{1}{a}, a]$ .

Let  $\Omega^l$  be the set of states which have exactly  $l$  loans with  $D$  and let  $\omega^l$  be a generic element of this set. Then the probability that the CDO defaults when no correlation is considered:

$$\sum_{l=\lceil \alpha n \rceil}^n \sum_{\omega^l \in \Omega^l} \eta^{NB}(\omega^l)$$

The worst case scenario among the scenarios determined by the correlation capacity  $a$  can be derived by using Theorem 1 as now stated:

**Lemma 3:** *The worst case scenario is that the CDO fails with probability*

$$\frac{a^2 \sum_{l=\lceil \alpha n \rceil}^n \sum_{\omega^l \in \Omega^l} \eta^{NB}(\omega^l)}{1 + (a^2 - 1) \sum_{l=\lceil \alpha n \rceil}^n \sum_{\omega^l \in \Omega^l} \eta^{NB}(\omega^l)}.$$

The result shows that the way to minimize beliefs about a set of states is a simple extension of how to minimize belief on one state. In other words, our analysis easily carries through for a combination of states.

Suppose that the rating agency chooses a triple A rating to the CDO if its probability of default is lower than some cutoff  $x$ . From the analysis above, we know that when  $n$  is small, the higher is  $a$  the more cautious will the rating agency be in granting a triple A rating. What happens when the CDO is complex, that is,  $n$  is large?

Let us consider the Naive-Bayes benchmark first. For a large  $n$ , one can approximate the Poisson Binomial distribution with a Poisson distribution with a mean  $\sum_{i=1}^n p_i$ .<sup>27</sup>

<sup>26</sup>See for example Wang et al (2009).

<sup>27</sup>See Hodges and Le Cam (1960).

**Example 4.** Suppose for that there is a share  $r$  of the assets with success rates  $p'$  and a share  $1 - r$  with success rate  $p''$ . Thus the mean for the Poisson approximation with  $l$  successes is  $n\mu$  where  $\mu = (rp' + (1 - r)p'')$ . If  $\mu > \alpha$ , there is a probability higher than a half that the CDO fails according to the independent approximation. Specifically, the probability that there are  $k$  or more defaults according to the Poisson distribution is  $1 - e^{-\mu n} \sum_{i=0}^{\lfloor l \rfloor} \frac{(\mu n)^i}{i!}$ . For this example, suppose that  $n = 100$ ,  $r = 0.6$  and  $p' = 0.8$  and  $p'' = 0.4$ , then  $\mu = 0.64$ . This implies that  $\Pr(\text{more than } \alpha n \text{ fail})$  equals 0.18 for  $\alpha = 0.7$ , 0.34 for  $\alpha = 0.66$ , 0.5 for  $\alpha = 0.64$ , 0.77 for  $\alpha = 0.6$ , and is close to 1 for  $\alpha = 0.5$  and values below.

More generally, we can show:

**Lemma 4:** *For large enough  $n$ , if  $\alpha > \mu$ , the CDO receives the highest rating for all  $a$ .*

The probability of each feasible state  $\omega$  does not become degenerate here, as opposed to the previous application. But the cumulative probability of many states together -which is the relevant one for the case of the CDO failing or not- does converge to be degenerate for some parameters, which therefore renders  $a$  immaterial. This implies that with complex assets composed of many assets, there are environments in which taking correlation into consideration (unless it is extreme as in the case of full correlation) will not change investors' behaviour. Even if the pessimists get their say in an organization, their recommendation would be to provide a high rating. In other words, the behavioural assumptions or the source of the correlation capacity is not important: in some environments, all would behave in the same manner of neglecting correlation when the asset is sufficiently complex. We therefore unravel a relation between complexity and correlation neglect.

We can use the above to derive some normative conclusions. Suppose that for all loans  $p_i = p$  for some  $p$ . Suppose first that the loans are completely independent. In that case, when  $n$  is small, some investors with a large  $a$  would be too cautious. For a large  $n$ , investors would behave efficiently as treating information as independent is the correct beliefs.

On the other hand when  $n$  is sufficiently large, even those with high  $a$  will behave as if they neglect correlation in some environments (when  $\alpha > p$ ). Suppose for example that the assets are fully correlated. Specifically, suppose the state  $\omega^*$  is drawn from  $\{D, ND\}$  using  $p$ , the probability of  $D$ , and all for all loans  $i$  we have  $\omega_i = \omega^*$ . Let  $p = 0.64$  (as is  $\mu$  in Example 4), which implies that the true probability of the CDO defaulting is 0.64. The efficient course of action is to award a triple A rating only if  $x < 0.64$  and not otherwise, irrespective of  $\alpha$ . However, whenever  $\alpha > 0.64$ , the CDO will be rated triple A for any  $x$  and for all  $a$ , inefficiently.

## 5 Extensions

We complete our analysis by providing two technical extensions to illustrate the robustness of Theorem 1.

## 5.1 Observing signals and marginals

In the analysis above we have assumed that the agent only observes the forecasts. This had allowed us to derive a set of rationalisable beliefs that is determined by the forecasts  $\mathbf{q}$  and not by the particulars of any information structure. An alternative assumption is that the agent also observes the marginal information structures of the sources or their signals.

In this Section we illustrate that relaxing these assumptions will not affect our qualitative results, that combining forecasts creates a set of rationalisable beliefs, that the set of combined forecasts is centered around the Naive-Bayes belief, and that it is convex. What does change is that the set of beliefs might depend on the particulars of these marginal information structures.

Consider for example the case explored in Example 2 from Section 2. Assume that there are two sources, (1 and 2) and consider for simplicity the case in which both have symmetric marginal information structures with binary signals ( $s^*$  and  $s^{**}$ ) about two possible realisations of the state (0 and 1). The agent then considers the set of possible symmetric joint information structures which is given by:

$$\begin{array}{cccccc} \omega = 0 & & s^* & & s^{**} & & \omega = 1 & & s^* & & s^{**} \\ s^* & & q(s^*|0) - q_0 & & q_0 & & s^* & & q(s^*|1) - q_1 & & q_1 \\ s^{**} & & q_0 & & 1 - q(s^*|0) - q_0 & & s^{**} & & q_1 & & 1 - q(s^*|1) - q_1 \end{array},$$

Note that even though the agent knows  $q(s|\omega)$ , ambiguity still exists because she does not know  $q_0$  and  $q_1$ . Suppose now that she observes the forecasts as well as  $q(s|\omega)$ , which is equivalent to observing the signals and marginals. Below, for expositional purposes, we focus on the case in which both information sources observed the signal  $s^*$ , i.e.,  $\mathbf{q} = (\frac{q(s^*|1)}{q(s^*|1)+q(s^*|0)}, \frac{q(s^*|1)}{q(s^*|1)+q(s^*|0)})$ .

We now characterise the set of rationalisable beliefs that can be derived from an information structure as above and which satisfies the ePMI constraints. As can be seen, our qualitative results hold in this case as well.

**Lemma 5:** *Given marginals  $q(s^*|\omega)$  and forecasts  $\mathbf{q} = (\frac{q(s^*|1)}{q(s^*|1)+q(s^*|0)}, \frac{q(s^*|1)}{q(s^*|1)+q(s^*|0)})$ , the set of rationalisable beliefs is: (i)  $C(a, \mathbf{q})$  as in Theorem 1 if  $\frac{q(s^*|1)}{1-q(s^*|1)} \leq a \leq \frac{1-q(s^*|0)}{q(s^*|0)}$ . (ii) Contained in  $C(a, \mathbf{q})$ , convex and contains the Naive-Bayes belief if  $a < \frac{q(s^*|1)}{1-q(s^*|1)}$  or  $\frac{1-q(s^*|0)}{q(s^*|0)} < a$ .*

## 5.2 Continuous distributions: The Morgenstern transformation

In this subsection we show that our analysis can be extended to continuous distributions. We use the Morgenstern transformation to derive a family of information structures starting from particular marginal information structures. This can then be useful in applications in



which continuous signal structures are more relevant.<sup>28</sup>

Suppose as above that the agent knows the marginal distributions as well as the signals of his information sources. Suppose that  $n = 1$ , and that the marginal distributions are symmetric,  $g(s^1|\omega)$  and  $g(s^2|\omega)$ , with PDFs  $G(s^1|\omega)$  and  $G(s^2|\omega)$  respectively, as in Example 3 in Section 2. We assume that the agent perceives the following family of joint information structures, constructed according to the Morgenstern transformation:

$$g(s^1, s^2|\omega) = [1 + \alpha_\omega(2G(s^1|\omega) - 1)(2G(s^2|\omega) - 1)]g(s^1|\omega)g(s^2|\omega).$$

In this family,  $\alpha_\omega > (<)0$  signifies positive (negative) correlation. For this family to hold, it has to be that  $|\alpha| \leq 1$ . Furthermore, to satisfy the ePMI constraints for some  $a$ , we also need that at any  $s^1, s^2, \omega$  :

$$\frac{1}{a} \leq [1 + \alpha_\omega(2G(s^1|\omega) - 1)(2G(s^2|\omega) - 1)] \leq a$$

which implies that  $\alpha$  is further constrained to be, at any  $\omega$  :

$$\frac{1}{a} - 1 \leq \alpha_\omega \leq 1 - \frac{1}{a}.$$

It is then easy to show that Theorem 1 holds as well.<sup>29</sup> The set of rationalisable beliefs given some  $s, s'$  is the set of all beliefs  $\eta(\cdot|s, s')$  satisfying:

$$\frac{\eta(\omega|s, s')}{\eta(\omega'|s, s')} = \frac{\gamma_\omega(s, s')g(s|\omega)g(s'|\omega)}{\gamma_{\omega'}(s, s')g(s|\omega')g(s'|\omega')} = \frac{\gamma_\omega(s, s')\eta^{NB}(\omega|s, s')}{\gamma_{\omega'}(s, s')\eta^{NB}(\omega'|s, s')},$$

for any  $\gamma_v(s, s') \in 1 + \alpha_v(2G(s|v) - 1)(2G(s'|v) - 1)$ , and  $\alpha_v \in [\frac{1}{a} - 1, 1 - \frac{1}{a}]$ , for  $v \in \{\omega, \omega'\}$ . Note that when  $\alpha_v = 0$  for all  $v$ , we have the Naive-Bayes benchmark as before.

## 6 Conclusion

We suggest a new framework to analyse how individuals combine forecasts when they have a bounded capacity for entertaining possible correlations in the joint information structure. We show that a larger capacity for correlation increases ambiguity and therefore conservative or cautious behaviour. On the other hand the level of information implicit in a Naive-Bayes interpretation of forecasts pushes individuals or organisations to be more confident and sometimes engage in risky behaviour. Moreover, when the Naive-Bayes interpretation is relatively precise, individuals are compelled to neglect correlation and thus in many “big data” environments we can expect the prevalence of correlation neglect.

<sup>28</sup>In Levy and Razin (2016) we use this transformation to analyze the effects of correlation capacity on common value auctions.

<sup>29</sup>The sufficiency part is as in Section 4. The necessity part follows directly from the assumption of the Morgenstern family of functions.

A next step in this research agenda will be to embed such decision making within strategic interaction or within an organisational design problem. For example, an interesting question arising from the description of the behaviour of financial institutions in the last decade is how the level of correlation capacity is determined. Moreover, it would be interesting to investigate the implications of heterogenous levels of capacity in market settings or in other strategic interactions.

As we have discussed in the introduction, combining forecasts is also relevant in many political environments. Poll aggregation is an obvious candidate for future analysis. In our applications above we considered two extreme cases; in one we focused on the correlation across fundamentals and in the other on the correlation across information sources. A simple electoral system, e.g., a referendum with a majority rule, can be captured by the latter case: The state of the world can be interpreted as a binary variable (e.g., Remain in the EU or Leave in the context of the EU referendum in Britain), and polls' predictions are clearly correlated to some degree as in many cases all use the same data set. A more complicated electoral system, such as the Electoral College for the US election, or the UK electoral system, are determined by a combination of outcomes in different regions (states in US states, constituencies in the UK). This case is similar to the CDO application in which correlation across the fundamentals is more pronounced, and a candidate wins if a some share of the regions votes in her favour. This latter case may be a more complicated environment for combining forecasts as the correlation across voters' preferences in different regions may be compounded with the correlation across polls. It may be an interesting question for future research to determine whether pollsters are less successful -perhaps due to the endogenous correlation neglect we had identified- in predicting outcomes in these more complicated environments.

## References

- [1] Al-Najjar, N. and J. Weinstein (2009). The Ambiguity Aversion Literature: A Critical Assessment. *Economics and Philosophy* 25, 249-284.
- [2] Avery, C. and P. Zemsky. (1998). Multidimensional Uncertainty and Herd Behavior in Financial Markets. *The American Economic Review*, Vol. 88, No. 4. pp. 724-748.
- [3] Baliga, S. E. Hanany, and P. Klibanoff (2013). Polarization and Ambiguity, *American Economic Review* 103(7): 3071–3083.
- [4] Banerjee, A. (1992). A Simple Model of Herd Behavior. *Quarterly Journal of Economics*, vol. 107, 797-817.
- [5] Bikhchandani, S. and S. Sharma (2000). Herd Behaviour in Financial Markets- A Review. IMF working paper WP/00/48.

- [6] Bikhchandani, S., D. Hirshleifer and I. Welch (1992), “A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades”, *Journal of Political Economy*, vol. 100, 992-1026.
- [7] Bohren, A. (2014), Informational Herding with Model Misspecification, mimeo.
- [8] Bose, S. and Renou, L. (2014). Mechanism Design with Ambiguous Communication Devices. *Econometrica* 82: 1853-1872.
- [9] Brunnermeier M.K. (2009). Deciphering the Liquidity and Credit Crunch 2007-2008. *Journal of Economic Perspectives*, 23, 77–100.
- [10] Budescu, D. and H. Yu (2007). Aggregation of Opinions Based on Correlated Cues and Advisors. *Journal of Behavioral Decision Making* 20: 153–177.
- [11] Church, K.W and P. Hanks (1991), Word association norms, mutual information, and lexicography. *Comput. Linguist.* 16 (1): 22–29.
- [12] Coval, J., J. Jurek, and E. Stafford (2009). The Economics of Structured Finance. *Journal of Economic Perspectives* 23, 3–25.
- [13] De Marzo PM, Vayanos D, Zwiebel J. (2003). Persuasion bias, social influence and unidimensional opinions. *Q. J. Econ.* 118:909–68.
- [14] Duffie, D., A. Eckner, G. Horel, and L. Saita. (2009). Frailty Correlated Default. *The Journal of Finance*, Vol. LXIV, No. 5, October.
- [15] Easley, D. and M. O’hara (2009), Ambiguity and Nonparticipation: The Role of Regulation, *Review of Financial Studies* 22, 1817-1843.
- [16] Ellis, A. and M. Piccione. Correlation Misperception in Choice. Forthcoming. *American Economic Review*.
- [17] Enke, B. and F. Zimmerman (2013). Correlation Neglect in Belief Formation. mimeo.
- [18] Epstein, L. and M. Schneider. (2007). Learning under Ambiguity. *Review of Economic Studies*. 74, 1275–1303.
- [19] Eyster, E. and M.Rabin. (2014). Extensive imitation is irrational and harmful. *The Quarterly Journal of Economics* [Internet]. 2014:qju021.
- [20] Eyster, E. and M. Rabin (2010). Naïve Herding in Rich-Information Settings. *American Economic Journal: Microeconomics*, 2(4): 221-43.
- [21] Eyster, E. and G. Weizsäcker (2011). Correlation Neglect in Financial Decision-Making. *Discussion Papers of DIW Berlin* 1104.

- [22] Gagnon-Bartsch, T. and M. Rabin (2015), Naive Social Learning, Mislearning, and Unlearning, mimeo.
- [23] Gilboa, I. and D. Schmeidler (1989), Maxmin expected utility with non-unique prior, *Journal of Mathematical Economics* 18, 141-153.
- [24] Glaeser, E. and Cass R. Sunstein (2009), Extremism and social learning.” *Journal of Legal Analysis*, Volume 1(1), 262 - 324.
- [25] Hanani, E. and P. Klibanoff (2007). Updating preferences with multiple priors. *Theoretical Economics* 2, 261-298.
- [26] Hodges J. and L. Le Cam (1960). The Poisson Approximation to the Poisson Binomial Distribution. *The Annals of Mathematical Statistics*. 737-740.
- [27] Jiang, J. and W. Tian (2016). Correlation Uncertainty, Heterogeneous Beliefs and Asset Prices. mimeo, University of North Carolina.
- [28] Kallir, I. and Sonsino, D. (2009). The Perception of Correlation in Investment Decisions. *Southern Economic Journal* 75 (4): 1045-66.
- [29] Levy, G. and R. Razin (2016). Common Value Auctions with Ambiguity over Information Structures and Correlation Capacities. mimeo.
- [30] Levy, G. and R. Razin. (2015a). Correlation Neglect, Voting Behaviour and Information Aggregation. *American Economic Review* 105: 1634-1645.
- [31] Levy, G. and R. Razin. (2015b). Does Polarisation of Opinions lead to Polarisation of Platforms? The Case of Correlation Neglect, *Quarterly Journal of Political Science*, Volume 10, Issue 3.
- [32] Lord, C., Lee Ross, and Mark R. Lepper. (1979). Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence. *Journal of Personality and Social Psychology* 37(11), 2098-2109.
- [33] Maines, L. (1990). The Effect of Forecast Redundancy on Judgments of a Consensus Forecast’s Expected Accuracy. *Journal of Accounting Research*, 28: 29-47
- [34] Maines, L. (1996). An Experimental Examination of Subjective Forecast Combination, *International Journal of Forecasting* 12 223-233.
- [35] Nelsen, R. (2006). *An Introduction to Copulas*. Springer.
- [36] Ortoleva, P. and E. Snowberg. (2015). Overconfidence in political economy. *American Economic Review*, 105: 504-535.

- [37] Schmid, F. and R. Schmidt (2007). Multivariate Extensions of Spearman’s rho and Related Statistics, *Statistics & Probability Letters*, Volume 77, Issue 4, Pages 407–416.
- [38] Seidenfeld, T. and Wasserman, L. (1993), Dilation for Sets of Probabilities. *Annals of Statistics*. 21 no. 3, 1139–1154.
- [39] Siniscalchi, M. (2011). Dynamic choice under ambiguity. *Theoretical Economics* 6, 379–421.
- [40] Sims, C. (2003), Implications of Rational Inattention, *Journal of Monetary Economics*, Volume 50, Issue 3, 665–690.
- [41] Sobel, J. (2014). On the relationship between individual and group decisions, *Theoretical Economics* 9, 163–185.
- [42] Vayanos, D. and J. Wang (2013). Market Liquidity|Theory and Empirical Evidence, in the *Handbook of the Economics of Finance*, edited by George Constantinides, Milton Harris, and Rene Stulz.
- [43] Wang, D., S.T. Rachev, F.J. Fabozzi (2009). Pricing Tranches of a CDO and a CDS Index: Recent Advances and Future Research. Part of the series *Contributions to Economics* pp 263-286.

## 7 Appendix

### 7.1 Appendix A

**Lemma 1a+Lemma1b:** In the text following Lemma 1b.■

#### **Proof of Theorem 1.**

We first consider  $n = 1$  and  $K > 1$ .

*Step 1:* Let  $\eta(\cdot) \in C(a, \mathbf{q})$ . Then there exists an information structure  $(S', q')$  with  $S' = \{s^*, s^{-*}\}^k$  which rationalises  $\eta(\cdot)$  and satisfies A1.

Assume that an information structure  $(S = \times_{j \in K} S^j, q(\mathbf{s}|\omega))$  rationalises  $\eta(\cdot)$ . Without loss of generality relabel signals so that the vector of signals that rationalises  $\eta(\omega)$  is  $(s^*, s^* \dots, s^*)$  so that  $\eta(\omega) = q(\omega|s^*, s^* \dots, s^*)$ . In addition we have that the following rationalizability and ePMI constraints are satisfied,

$$\forall j \in K \text{ and } \forall \omega \in \Omega, \quad q^j(\omega) = q^j(\omega|s^*)$$

$$\forall \mathbf{s} = (s_1, \dots, s_k) \in \times_{j \in K} S^j \text{ and } \forall \omega \in \Omega, \quad \frac{1}{a} \leq \frac{q(\mathbf{s}|\omega)}{\prod_{j \in K} q^j(s^j|\omega)} \leq a.$$

Construct the new information structure  $(S', q'(\cdot|\omega))$  by keeping the same distribution over signals as in  $(S, q)$ , while keeping the label  $s^*$  and bundling all possible signals  $s \neq s^*$  under one signal  $s^{-*}$ . In particular,  $\forall \omega \in \Omega$ ,

$$\begin{aligned} q'(s^*, \dots, s^*|\omega) &= q(s^*, \dots, s^*|\omega) \\ q'(s^{-*}, s^*, \dots, s^*|\omega) &= \sum_{s \in S^1/\{s^*\}} q(s, s^*, \dots, s^*|\omega), \end{aligned}$$

and so on. Note that  $(S', q')$  rationalises  $\eta(\cdot)$  by definition.

It remains to show that the ePMI constraints hold for  $(S', q')$  so that it satisfies A1. Note first that the ePMI constraint for  $(s^*, \dots, s^*)$  holds by definition of  $(S', q')$ . Consider any other profile of signals  $\mathbf{s} \in \{s^*, s^{-*}\}^k$ . The ePMI constraint for  $\mathbf{s}$  can be expressed in terms of the information structure  $(S, q)$  as  $\frac{\sum_{l=1}^m c_l}{\sum_{l=1}^m c'_l}$  where  $c_l = q(\mathbf{s}_l|\omega)$  for some  $\mathbf{s}_l = (s_l^1, \dots, s_l^k) \in S$  where we sum over all  $\mathbf{s}^l$  that compose  $\mathbf{s}$ , and  $c'_l = \prod_{j \in K} q^j(s_l^j|\omega)$ . But as the original ePMI constraints hold, this also implies that  $\frac{1}{a} \leq \frac{\sum_{l=1}^m c_l}{\sum_{l=1}^m c'_l} \leq a$ . Thus the ePMI constraints are satisfied also for  $(S', q')$ . ■

Wlog assume that the agent rationalizes the set of posteriors she observes by believing that all sources have received the signal  $s^*$ . For any  $v \in \Omega$ , let  $\alpha_v = \Pr(\text{all receive } s^*|v)$  and let  $\delta_v^i = \Pr(i \text{ receives } s^*|v)$ .

*Step 2: Suppose  $n=1$  and  $K>1$ . For any  $\eta(\omega)$  that satisfies the necessary condition in the Theorem, there exists an information structure that satisfies A1 and rationalizes this belief.*

Take any vector  $(\lambda_\omega)_{\omega \in \Omega}$  that satisfies  $\frac{1}{a} \leq \lambda_\omega \leq a$  for any realisation of  $\omega$  and consider the belief

$$\eta(\omega) = \frac{\lambda_\omega \frac{1}{p(\omega)^{k-1}} \prod_{j \in K} q^j(\omega)}{\sum_{v \in \Omega} \lambda_v \frac{1}{p(v)^{k-1}} \prod_{j \in K} q^j(v)}.$$

Using this vector  $(\lambda_\omega)_{\omega \in \Omega}$  we now construct an information structure that will satisfy all ePMI constraints and the rationalisability constraints, and will rationalise the belief  $\eta(\omega)$ .

Let  $\alpha_\omega = \lambda_\omega \prod_{j \in K} \delta_\omega^j$  and let  $\delta_\omega^j = \varepsilon \frac{q^j(\omega)}{p(\omega)}$ . This implies that this information structure

generates the belief as desired as  $\eta(\omega) = \frac{p(\omega)\alpha_\omega}{\sum_{v \in \Omega} p(v)\alpha_v} = \frac{\lambda_\omega \frac{1}{p(\omega)^{k-1}} \prod_{j \in K} q^j(\omega)}{\sum_{v \in \Omega} \lambda_v \frac{1}{p(v)^{k-1}} \prod_{j \in K} q^j(v)}$ .

Note that  $\frac{p(\omega)\delta_\omega^j}{\sum_{v \in \Omega} p(v)\delta_v^j} = \frac{q^j(\omega)}{\sum_{v \in \Omega} q^j(v)} = q^j(\omega)$  which implies that the posterior beliefs of all individuals are rationalized.

We now specify the joint distribution over signals, making sure that all the ePMI constraints are satisfied. For all  $\omega \in \Omega$ , set the joint probability of each event in which two or more sources receive  $s^*$ , but not when all sources receive  $s^*$ , to satisfy independence. For example, the probability that all  $m$  sources in the set  $M$  and only these individuals receive

$s^*$  in state  $\omega$ , for  $1 < m < k$ , is  $\prod_{j \in M} \delta_\omega^j \prod_{j \in K/M} (1 - \delta_\omega^j)$ . Thus for all these cases the ePMI constraints are satisfied.

At any state, we then need to verify the ePMI constraints in the following events: when one source exactly had received  $s^*$ , or when all received  $s^{-*}$ . Let us focus on some realisation  $\omega$ . Consider first the event in which only one source had received  $s^*$ .

$$\Pr(s^j = s^*, \text{ all others receive } s^{-*} | \omega) = \varepsilon \frac{q^j(\omega)}{p(\omega)} - \alpha_\omega - \varepsilon \frac{q^j(\omega)}{p(\omega)} \left( \sum_{\substack{M \subset K/\{j\} \\ |M| \geq 1}} \prod_{i \in M} \delta_\omega^i \prod_{l \in K/M \cup \{j\}} (1 - \delta_\omega^l) \right)$$

The ePMI is:

$$\begin{aligned} & \frac{\varepsilon \frac{q^j(\omega)}{p(\omega)} - \alpha_\omega - \varepsilon \frac{q^j(\omega)}{p(\omega)} \left( \sum_{\substack{M \subset K/\{j\} \\ |M| \geq 1}} \prod_{i \in M} \delta_\omega^i \prod_{l \in K/M \cup \{j\}} (1 - \delta_\omega^l) \right)}{\varepsilon \frac{q^j(\omega)}{p(\omega)} \prod_{l \neq j} (1 - \varepsilon \frac{q^l(\omega)}{p(\omega)})} \\ = & \frac{1 - \lambda_\omega \prod_{l \in K/j} \left( \varepsilon \frac{q^l(\omega)}{p(\omega)} \right) - \left( \sum_{\substack{M \subset K/\{j\} \\ |M| \geq 1}} \prod_{i \in M} \delta_\omega^i \prod_{l \in K/M \cup \{j\}} (1 - \delta_\omega^l) \right)}{\prod_{l \neq j} (1 - \varepsilon \frac{q^l(\omega)}{p(\omega)})} \rightarrow_{\varepsilon \rightarrow 0} 1, \end{aligned}$$

as for all  $k$ ,  $\delta_\omega^k$  goes to 0 with  $\varepsilon$ . Thus, the ePMI can be made smaller than  $a$  and greater than  $\frac{1}{a}$ , if  $\varepsilon$  is small enough.

Consider now the event that all sources had received  $s^{-*}$  in state  $\omega$ :

$$\begin{aligned} \Pr(\text{all received signal } s^{-*} | \omega) &= \left(1 - \varepsilon \frac{q^j(\omega)}{p(\omega)}\right) - \left(1 - \varepsilon \frac{q^j(\omega)}{p(\omega)}\right) \left( \sum_{\substack{M \subset K/\{j\} \\ |M| \geq 2}} \prod_{i \in M} \delta_\omega^i \prod_{l \in K/M \cup \{j\}} (1 - \delta_\omega^l) \right) \\ &\quad - (k-1) \left( \varepsilon \frac{q^j(\omega)}{p(\omega)} - \alpha_\omega - \varepsilon \frac{q^j(\omega)}{p(\omega)} \left( \sum_{\substack{M \subset K/\{j\} \\ |M| \geq 1}} \prod_{i \in M} \delta_\omega^i \prod_{l \in K/M \cup \{j\}} (1 - \delta_\omega^l) \right) \right), \end{aligned}$$

where here we subtract all the events in which two or more received  $s^*$  (but at most  $k-1$ ), and the  $k-1$  events in which just one player had received  $s^*$  which we had described above.

The ePMI is:

$$\begin{aligned}
& \frac{(1-\varepsilon \frac{q^j(\omega)}{p(\omega)})}{(1-\varepsilon \frac{q^j(\omega)}{p(\omega)}) \prod_{l \neq j} (1-\varepsilon \frac{q^l(\omega)}{p(\omega)})} \\
& \frac{(1-\varepsilon \frac{q^j(\omega)}{p(\omega)}) (\sum_{\substack{M \subset K / \{j\} \\ |M| \geq 2}} \prod_{i \in M} \delta_\omega^i \prod_{l \in K/M \cup \{j\}} (1-\delta_\omega^l)) - (k-1) (\varepsilon \frac{q^j(\omega)}{p(\omega)} - \alpha_\omega - \varepsilon \frac{q^j(\omega)}{p(\omega)}) (\sum_{\substack{M \subset K / \{j\} \\ |M| \geq 1}} \prod_{i \in M} \delta_\omega^i \prod_{l \in K/M \cup \{j\}} (1-\delta_\omega^l))}{(1-\varepsilon \frac{q^j(\omega)}{p(\omega)}) \prod_{l \neq j} (1-\varepsilon \frac{q^l(\omega)}{p(\omega)})} \\
& = \frac{1 - (\sum_{\substack{M \subset K / \{j\} \\ |M| \geq 2}} \prod_{i \in M} \delta_\omega^i \prod_{l \in K/M \cup \{j\}} (1-\delta_\omega^l))}{\prod_{l \neq j} (1-\varepsilon \frac{q^l(\omega)}{p(\omega)})} \\
& - \varepsilon \frac{(k-1) (\frac{q^j(\omega)}{p(\omega)} - \lambda_\omega \varepsilon^{k-1} \prod_{i \in K} (\frac{q^i(\omega)}{p(\omega)}) - \frac{q^j(\omega)}{p(\omega)}) (\sum_{\substack{M \subset K / \{j\} \\ |M| \geq 1}} \prod_{i \in M} \delta_\omega^i \prod_{l \in K/M \cup \{j\}} (1-\delta_\omega^l))}{(1-\varepsilon \frac{q^j(\omega)}{p(\omega)}) \prod_{l \neq j} (1-\varepsilon \frac{q^l(\omega)}{p(\omega)})} \\
& \rightarrow \varepsilon \rightarrow 0 \mathbf{1}.
\end{aligned}$$

which again can be made smaller than  $a$  and larger than  $\frac{1}{a}$  for low enough  $\varepsilon$ . Thus all constraints in state  $\omega$  can be satisfied. ■

*Step 3: Suppose now that  $n > 1$ . For any  $\eta(\cdot)$  that satisfies the necessary condition in the Theorem, there exists an information structure that satisfies A1 and rationalizes this belief.*

Consider the belief  $\eta(\omega) = \frac{\lambda_\omega \eta^{NB}(\omega)}{\sum_v \lambda_v \eta^{NB}(v)}$ . Let  $q(s, \omega) = \prod_{j \in K} q_i^j(s_i^j | \omega_i)$ . Let  $q_i^j(s_i^j | \omega_i) = \frac{q_i^j(\omega_i)}{p_i(\omega_i)}$  and let  $p(\omega) = \lambda_\omega \prod_{i \in N} p_i(\omega_i)$ . The ePMI constraints are satisfied as well as the rationalisability constraints as  $\frac{p_i(\omega_i) q_i^j(s_i^j | \omega_i)}{\sum_{v_i} p_i(v_i) q_i^j(s_i^j | v_i)} = \frac{q_i^j(\omega_i)}{\sum_{v_i} q_i^j(v_i)} = q_i^j(\omega_i)$ . Moreover the belief can be generated by  $\eta(\omega) = \frac{p(\omega) \prod_{i \in N} \prod_{j \in K_i} q_i^j(s_i^j | \omega_i)}{\sum_v p(v) \prod_{i \in N} \prod_{j \in K_i} q_i^j(s_i^j | v_i)} = \frac{p(\omega) \prod_{i \in N} \prod_{j \in K_i} q_i^j(s_i^j | \omega_i)}{\sum_v p(v) \prod_{i \in N} \prod_{j \in K_i} q_i^j(s_i^j | v_i)} = \frac{\lambda_\omega \eta^{NB}(\omega)}{\sum_v \lambda_v \eta^{NB}(v)}$  as desired. ■

*Step 4:  $C(a, \mathbf{q})$  is compact and convex.*

Compactness comes from the proof in the text and the previous steps. To prove convexity consider two beliefs  $\eta$  and  $\eta'$  that are in  $C(a, \mathbf{q})$ . Note that from the above a belief  $\eta(\cdot)$  is in  $C(a, \mathbf{q})$  if and only if for any  $v, \omega \in \Omega$  we have,

$$\frac{\eta(\omega)}{\eta(v)} = \frac{\lambda_\omega \eta^{NB}(\omega)}{\lambda_v \eta^{NB}(v)}.$$

Thus all likelihood ratios satisfy,

$$\frac{1}{a^2} \frac{\eta^{NB}(\omega)}{\eta^{NB}(v)} \leq \frac{\eta(\omega)}{\eta(v)} \leq a^2 \frac{\eta^{NB}(\omega)}{\eta^{NB}(v)}. \quad (3)$$

To prove convexity we show that we can find a vector  $\lambda^\beta$  with elements between  $\frac{1}{a}$  and  $a$  that spans  $\beta\eta + (1-\beta)\eta'$ . It will be enough to show that  $\beta\eta + (1-\beta)\eta'$  has likelihood ratios in the bounds in (3). Note that  $\eta, \eta'$  satisfy



$$\frac{1^2 \eta^{NB}(\boldsymbol{\omega})}{a \eta^{NB}(\mathbf{v})} \leq \frac{\eta(\boldsymbol{\omega})}{\eta(\mathbf{v})} \leq a^2 \frac{\eta^{NB}(\boldsymbol{\omega})}{\eta^{NB}(\mathbf{v})}, \quad \frac{1^2 \eta^{NB}(\boldsymbol{\omega})}{a \eta^{NB}(\mathbf{v})} \leq \frac{\eta'(\boldsymbol{\omega})}{\eta'(\mathbf{v})} \leq a^2 \frac{\eta^{NB}(\boldsymbol{\omega})}{\eta^{NB}(\mathbf{v})},$$

we have that:

$$\frac{\beta \eta(\boldsymbol{\omega}) + (1 - \beta) \eta'(\boldsymbol{\omega})}{\beta \eta(\mathbf{v}) + (1 - \beta) \eta'(\mathbf{v})} \leq \frac{\beta \eta(\mathbf{v}) + (1 - \beta) \eta'(\mathbf{v})}{\beta \eta(\mathbf{v}) + (1 - \beta) \eta'(\mathbf{v})} a^2 \frac{\eta^{NB}(\boldsymbol{\omega})}{\eta^{NB}(\mathbf{v})} = a^2 \frac{\eta^{NB}(\boldsymbol{\omega})}{\eta^{NB}(\mathbf{v})},$$

and similarly that,

$$\frac{\beta \eta(\boldsymbol{\omega}) + (1 - \beta) \eta'(\boldsymbol{\omega})}{\beta \eta(\mathbf{v}) + (1 - \beta) \eta'(\mathbf{v})} \geq \frac{1^2 \eta^{NB}(\boldsymbol{\omega})}{a \eta^{NB}(\mathbf{v})}.$$

So there must exist  $\boldsymbol{\lambda}^\beta$  that spans  $\beta \eta + (1 - \beta) \eta'$ . ■

**Proof of Lemma 2:** (i) The proof follows from the construction in Theorem 1, Corollary 1 and maxmin preferences. These imply that an individual  $i$  who has a lower perception of correlation than an individual  $j$ , will choose to invest according to a higher belief about state 1 and hence will invest more in risky asset. (ii) The proof follows from the construction in Theorem 1. Fix  $K, N$  and  $q$ , as  $a$  goes to infinity, the set  $C(a, \mathbf{q})$  converges to span all possible beliefs. Therefore there is a  $\gamma > 1$  such that if  $\gamma < a$ , each investor will have a minimum belief that is lower than his  $q^i(1)$  and hence will experience a cautious shift. (iii) This is explained in the text. ■

**Proof of Lemma 3:** First note that our results extend to a combination of states. That is, we know that the maximum belief in the set

$$\eta(\boldsymbol{\omega}) = \frac{\lambda_{\boldsymbol{\omega}} \eta^{NB}(\boldsymbol{\omega})}{\lambda_{\boldsymbol{\omega}} \eta^{NB}(\boldsymbol{\omega}) + \sum_{\boldsymbol{\omega}' \neq \boldsymbol{\omega}} \lambda_{\boldsymbol{\omega}'} \eta^{NB}(\boldsymbol{\omega}')}$$

where  $\lambda_{\mathbf{v}} \in [1/a, a]$ , is attained when  $\lambda_{\boldsymbol{\omega}} = a$  and  $\lambda_{\boldsymbol{\omega}'} = 1/a$  for all other  $\boldsymbol{\omega}'$ . But also the maximum in the set

$$\eta(\boldsymbol{\omega}) + \eta(\boldsymbol{\omega}') = \frac{\lambda_{\boldsymbol{\omega}} \eta^{NB}(\boldsymbol{\omega}) + \lambda_{\boldsymbol{\omega}'} \eta^{NB}(\boldsymbol{\omega}')}{\lambda_{\boldsymbol{\omega}} \eta^{NB}(\boldsymbol{\omega}) + \lambda_{\boldsymbol{\omega}'} \eta^{NB}(\boldsymbol{\omega}') + \sum_{\mathbf{v} \neq \boldsymbol{\omega}, \boldsymbol{\omega}'} \lambda_{\mathbf{v}} \eta^{NB}(\mathbf{v})},$$

by taking derivatives w.r.t. the  $\lambda$ 's, is attained when  $\lambda_{\boldsymbol{\omega}}, \lambda_{\boldsymbol{\omega}'} = a$  and  $\lambda_{\mathbf{v}} = 1/a$  for all others.

Thus the worst case scenario is the highest belief that the CDO fails meaning:

$$\frac{a \sum_{l=\lceil \alpha n \rceil}^n \sum_{\boldsymbol{\omega}^l \in \Omega^l} \eta^{NB}(\boldsymbol{\omega}^l)}{a \sum_{l=\lceil \alpha n \rceil}^n \sum_{\boldsymbol{\omega}^l \in \Omega^l} \eta^{NB}(\boldsymbol{\omega}^l) + \frac{1}{a} (1 - \sum_{l=\lceil \alpha n \rceil}^n \sum_{\boldsymbol{\omega}^l \in \Omega^l} \eta^{NB}(\boldsymbol{\omega}^l))}$$

which equals the formulation in the text. ■

**Proof of Lemma 4:** By the approximation and from Lemma 3, we know that the worst case scenario is

$$\frac{a(1 - e^{-\mu n} \sum_{i=0}^{\lfloor \alpha n \rfloor} \frac{(\mu n)^i}{i!})}{a(1 - e^{-\mu n} \sum_{i=0}^{\lfloor \alpha n \rfloor} \frac{(\mu n)^i}{i!}) + (1/a)(e^{-\mu n} \sum_{i=0}^{\lfloor \alpha n \rfloor} \frac{(\mu n)^i}{i!})}$$

But note that for any  $\alpha > \mu$  we have  $\lim_{n \rightarrow \infty} (1 - e^{-\mu n} \sum_{i=0}^{\lfloor \alpha n \rfloor} \frac{(\mu n)^i}{i!}) = 0$ , implying that

$$\lim_{n \rightarrow \infty} \frac{a(1 - e^{-\mu n} \sum_{i=0}^{\lfloor \alpha n \rfloor} \frac{(\mu n)^i}{i!})}{a(1 - e^{-\mu n} \sum_{i=0}^{\lfloor \alpha n \rfloor} \frac{(\mu n)^i}{i!}) + (1/a)(e^{-\mu n} \sum_{i=0}^{\lfloor \alpha n \rfloor} \frac{(\mu n)^i}{i!})} = 0$$

Therefore, for any  $\alpha > \mu$  we have that, for all  $x$  and for all  $a$ , the CDO is deemed safe. ■

**Proof of Lemma 5:** The ePMI constraints are:

$$\begin{aligned} \frac{1}{a} &\leq \frac{q(s^*|\omega) - q_\omega}{q(s^*|\omega)^2} \leq a \\ \frac{1}{a} &\leq \frac{q_\omega}{q(s^*|\omega)(1 - q(s^*|\omega))} \leq a \\ \frac{1}{a} &\leq \frac{1 - q(s^*|\omega) - q_\omega}{(1 - q(s^*|\omega))^2} \leq a \end{aligned}$$

First note that the Naive-Bayes belief satisfies all the constraints. Note that the belief that the state is one is  $\frac{\alpha_1}{\alpha_0 + \alpha_1}$ . We proceed by characterising the highest and lowest values we can get for  $\alpha_\omega$ .

From the first ePMI constraints we have that:  $aq(s^*|\omega)^2 \geq q(s^*|\omega)^2 - q_\omega \geq \frac{1}{a}q(s^*|\omega)^2$ . Note the third ePMI constraints above do not bind at the extremes of the above inequalities:

$$\frac{1}{a} \leq \frac{1 - 2q(s^*|\omega) + \frac{1}{a}q(s^*|\omega)^2}{(1 - q(s^*|\omega))^2} < \frac{1 - 2q(s^*|\omega) + q(s^*|\omega)^2}{(1 - q(s^*|\omega))^2} = 1 < a,$$

where the LHS inequality is derived from

$$\frac{1}{a} \leq \frac{1 - 2q(s^*|\omega) + \frac{1}{a}q(s^*|\omega)^2}{(1 - q(s^*|\omega))^2} \Leftrightarrow (1 - 2q(s^*|\omega)) \leq a(1 - 2q(s^*|\omega)) \Leftrightarrow 1 \leq a.$$

Similarly,

$$\frac{1}{a} \leq 1 = \frac{1 - 2q(s^*|\omega) + q(s^*|\omega)^2}{(1 - q(s^*|\omega))^2} \leq \frac{1 - 2q(s^*|\omega) + aq(s^*|\omega)^2}{(1 - q(s^*|\omega))^2} \leq a,$$

where the RHS inequality is derived from

$$\frac{1 - 2q(s^*|\omega) + aq(s^*|\omega)^2}{(1 - q(s^*|\omega))^2} \leq a \Leftrightarrow 1 - 2q(s^*|\omega) \leq a(1 - 2q(s^*|\omega)) \Leftrightarrow 1 \leq a.$$

So the only constraints left are  $\frac{1}{a} \leq \frac{q_\omega}{q(s^*|\omega)(1-q(s^*|\omega))} \leq a$ . For this the extremes could matter as while  $\frac{1}{a} \leq \frac{q(s^*|\omega) - \frac{1}{a}q(s^*|\omega)^2}{q(s^*|\omega)(1-q(s^*|\omega))}$  is satisfied as  $\frac{q(s^*|\omega) - \frac{1}{a}q(s^*|\omega)^2}{q(s^*|\omega)(1-q(s^*|\omega))} > \frac{q(s^*|\omega) - q(s^*|\omega)^2}{q(s^*|\omega)(1-q(s^*|\omega))} = 1 \geq \frac{1}{a}$ , the other side has  $\frac{q(s^*|\omega) - \frac{1}{a}q(s^*|\omega)^2}{q(s^*|\omega)(1-q(s^*|\omega))} \leq a \Leftrightarrow \frac{q(s^*|\omega)}{(1-q(s^*|\omega))} \leq a$  and similarly, for the other extreme, we will need  $\frac{1}{a} \leq \frac{q(s^*|\omega) - aq(s^*|\omega)^2}{q(s^*|\omega)(1-q(s^*|\omega))} \Leftrightarrow q(s^*|\omega)(a+1) \leq 1 \Leftrightarrow a \leq \frac{1-q(s^*|\omega)}{q(s^*|\omega)}$ . ■

## 7.2 Appendix B: Other results

### 7.2.1 Pointwise mutual information and concordance

**Proposition B1:** *Assume that there are two information sources,  $k$  and  $j$ . There is a  $0 < \bar{\rho} < 1$  such that any joint information structure that satisfies A1 has a Spearman's  $\rho$  (Kendal's  $\tau$ ) in  $[-\bar{\rho}, \bar{\rho}]$ .*

**Proof of Proposition B1:** The bounds on the ePMI imply that there is an  $\varepsilon$  such that  $\frac{q(s^k, s^j|\omega)}{q^k(s^k|\omega)q^j(s^j|\omega)} \in [1 - \varepsilon, 1 + \varepsilon]$ . This implies that  $|q(s^k, s^j|\omega) - q^k(s^k|\omega)q^j(s^j|\omega)| \leq \varepsilon q^k(s^k|\omega)q^j(s^j|\omega)$ . Summing up over all  $(s^k, s^j)$  and given  $x, y$  we get that  $|Q(x, y|\omega) - Q^k(x|\omega)Q^j(y|\omega)| \leq \varepsilon Q^k(x|\omega)Q^j(y|\omega) \leq \varepsilon$ . This implies that the distance between the copula of any such information structure to the product copula is bounded by  $\varepsilon$ .

Among all such information structures, take the supremum according to the highest copula. That information structure has a Spearman's  $\rho$  (Kendall's  $\tau$ ) that is strictly smaller than 1 (See Theorem 5.9.6 and Theorem 5.1.3 in Nelsen 2006).

Among all such information structures, take the infimum according to the lowest copula. That information structure has a Spearman's  $\rho$  (or Kendall's  $\tau$ ) that is strictly larger than -1 (See Theorem 5.9.6 and Theorem 5.1.3 in Nelsen 2006).

By Theorem 5.1.9 in Nelsen (2006), any other information structure will have a Spearman's  $\rho$  (Kendall's  $\tau$ ) in between the two copulas above. ■

### 7.2.2 Agents with private information

In the application in Section 4.1, each agent receives a signal and generates a prediction. There are two subtleties to consider in order to extend the model described in Section 2. First, when the agent receives a signal, knows his marginal distribution, and updates his belief, we need to show that he ends up with a unique rationalised belief even though he can imagine many joint information structures. This we do in Lemma B1. Second, to extend our model directly from there, we need to assume that individuals forget their marginals and signals when they combine forecasts, so the only information they have is the vector  $\mathbf{q}$ . In this case we are exactly in the same model as in Section 2. But note that if not, our results still hold. Specifically, in Lemma 5 we show that the results extend to the case in which the agent knows the marginal distributions and signals.

**Lemma B1:** *Suppose that each agent receives a signal and knows his marginal . Then*

each agent has a unique posterior. That is, given an observation of some  $s' \in S^j$ , individual  $j$  updates his belief to  $q^j(\omega|s') = \frac{p(\omega)q^j(s'|\omega)}{\sum_{v \in \Omega} p(v)q^j(s'|v)}$ .

**Proof:** Individual  $j$  observes  $s' \in S^j$  and considers all joint information structures which have a marginal information structure that accords with his own. That is, all  $(\times_{l=1}^k \hat{S}^l, \hat{q}(\mathbf{s}, \omega))$  for which  $\sum_{\mathbf{s}^{-j} \in \times_{j \neq l} \hat{S}^l} \hat{q}(s', \mathbf{s}^{-j}|\omega) = q^j(s'|\omega)$  for all  $\omega$ . For any such joint information structure  $(\times_{l=1}^k \hat{S}^l, \hat{q}(\mathbf{s}, \omega))$ , we generate the posterior belief about state  $\omega$  as

$$\hat{q}^j(\omega|s') = \frac{\sum_{\mathbf{s}^{-j} \in \times_{j \neq l} \hat{S}^l} p(\omega)\hat{q}(s', \mathbf{s}^{-j}|\omega)}{\sum_{v \in \Omega} \sum_{\mathbf{s}^{-j} \in \times_{j \neq l} \hat{S}^l} p(v)\hat{q}(s', \mathbf{s}^{-j}|v)} = \frac{p(\omega)q^j(s'|\omega)}{\sum_{v \in \Omega} p(v)q^j(s'|v)}$$

for all  $\omega$ . ■

Interestingly, the above implies then that ambiguity over joint information structures implies ambiguity over the state of the world but only following exposure to multiple information sources. This is related to the notion of dilation introduced in Seidenfeld and Wasserman (1993). Seidenfeld and Wasserman (1993) focus on lower and upper probability bounds for probability events. Dilation is defined as a situation in which the probability bounds of an event A are strictly within the probability bounds for the event in which A is conditional on B. When we compare an individual's private belief to the set of beliefs he gains after observing multiple sources, sometimes dilation occurs.<sup>30</sup>

---

<sup>30</sup>Similar dilation effects also occur in Bose and Renou (2014) who study how principals can use ambiguous mechanisms to implement social welfare functions that are not attainable under unambiguous mechanisms. Epstein and Schneider (2007) analyze a learning process in which an agent receives a stream of signals over time. The function mapping the state of the world and the signals is not constant over time and thus additional signals create ambiguity. Our analysis differs in an important way: the signals that the different agents receive in our model may be correlated whereas in their model, the stream of signals one agent receives is known to be independent.