Combining Forecasts in the Presence of Ambiguity over Correlation Structures

Gilat Levy and Ronny Razin, LSE[1]

Abstract: We suggest a framework to analyse how sophisticated decision makers combine multiple sources of information to form predictions. In particular, we focus on situations in which: (i) Decision makers understand each information source in isolation but are uncertain about the correlation between the sources; (ii) Decision makers consider a range of bounded correlation scenarios to yield a set of possible predictions; (iii) Decision makers face ambiguity in relation to the set of predictions they consider. We measure the bound on correlation scenarios by using the notion of pointwise mutual information. We show that the set of predictions the decision makers considers is completely characterised by two parameters: the Naïve-Bayes interpretation of forecasts (correlation neglect), and the bound on the correlation between information sources. The analysis yields two countervailing effects on behaviour. First, when the Naïve-Bayes interpretation of information is relatively precise, it can induce risky behaviour, irrespective of what correlation scenario is chosen. Second, a higher correlation bound creates more uncertainty and therefore potentially more conservative behaviour. We show how this trade-off affects behaviour in different applications, including financial investments, group decision making and CDO ratings. For the latter, we show that when faced with complex assets, decision makers are likely to behave in ways that are consistent with complete correlation neglect.

[1] 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, Cornell, PSE, ESSET 2015, Manchester University, University of Bonn, LBS, Ecole Polytechnique, Bristol University, University of Bath, 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. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 681579.

1 Introduction

When confronted with multiple forecasts, we often have a better understanding of each forecast separately than we do of how the sources relate to one another. This is apparent in many situations when experts or organisations make predictions. In the finance literature this has been long recognised.[2] Jiang and Tian (2016) point to several problems in estimating correlation, including the lack of sufficient market data, instabilities in the correlation process and the increasingly interconnected market patterns. The US financial crisis inquiry (FCIC) report from 2011 cites the acknowledgment of the rating agency Moody's that "In the absence of meaningful default data, it is impossible to develop empirical default correlation measures based on actual observations of defaults."

In this paper we suggest a framework to model how sophisticated individuals combine forecasts in complex environments. In particular, we focus on situations in which: (i) Decision makers understand each information source in isolation but are uncertain about the correlation between the sources; (ii) Decision makers consider a range of bounded correlation scenarios to yield a set of possible predictions; (iii) Decision makers face ambiguity in relation to the set of predictions they consider.

In particular, we consider an environment in which an agent observes forecasts about a potentially multidimensional state of the world, $\boldsymbol{\omega} \in \Omega^n$. For each element ω_i in $\boldsymbol{\omega}$, the agent observes possibly multiple forecasts, each a probability distribution over Ω. To combine the multiple forecasts into a prediction about $\boldsymbol{\omega}$, the agent considers a set of possible joint information structures that could have yielded these forecasts. We allow for two types of correlation in the consideration set of the agent: across the fundamentals (the elements in $\boldsymbol{\omega}$), and across the predictions (e.g., due to biases in polling techniques). 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 $\boldsymbol{\omega}$ that is the focus of our analysis. For example, if the decision maker only considers joint information structures which satisfy (conditional) independence across forecasts, then the unique prediction that arises is the Naïve-Bayes (NB) belief.

Our main modeling assumption is to use a bound on the *pointwise mutual information* (PMI) of information structures as the bound on the correlation scenarios the decision maker

[2] This is the motivation behind papers such as Duffie et al (2009).

considers. PMI relates to the distance between the joint distribution and the independent benchmark, that is, the multiplication of the marginal distributions. The higher is the bound on the PMI, the more correlation levels can be considered. As we show, modelling the perceptions of individuals about correlation in this way is general, distribution free, and allows us to complete the model by using attitudes towards ambiguity as a model of decision making when decision makers face multiple predictions.

We characterise the set of predictions of a decision maker who considers scenarios with bounded correlation structures. We show that the set of predictions is convex and compact, and is monotonic (set-wise) in the PMI bound. Moreover, it can be fully characterised by two sufficient statistics: The PMI-bound and the Naïve-Bayes belief that assumes independence.

The above results allow us to analyze the ambiguity of decision makers vis a vis the set of predictions they have generated. In particular, the convexity of the set of predictions allows us to treat them as a set of priors. Moreover, the (set) monotonicity implies a metric by which ambiguity increases as the PMI bound increases. Specifically, the model implies that larger ambiguity over correlation structures can translate to larger ambiguity over the state of the world.

In contrast to the effect of a higher PMI bound, we show that as the NB prediction becomes more informative the set of predictions shrinks. When the NB prediction converges to a degenerate belief, all decision makers, whatever their preferences over ambiguity or their PMI bounds, will make the same decisions. In particular, fixing the PMI-bound when the NB belief becomes degenerate, the set of predictions shrinks and converges to the NB prediction. In an application to the evaluation of financial assets we formalise the notion that the complexity of securities distorts its evaluation towards correlation neglect. Focusing on CDOs, we show that as the number of individual mortgages in a CDO increases, its evaluation becomes highly dependent on the NB belief. Thus, the evaluation of complex assets might suffer from complete correlation neglect even when experts allow for a wide family of correlation scenarios.

Our model therefore highlights an intuitive relation between the set of correlation structures the decision maker considers and her confidence in the decision she takes. First, when the NB belief is relatively precise, the decision maker behaves as if she completely neglects correlation, which, as already explored in the literature, implies more extreme beliefs.[3] Sec-

[3]This arises -with standard information structures such as the normal distribution- in Ortoleva and

ond, correlation will affect confidence through a Knightian notion of uncertainty. A decision maker who is ambiguity averse and who entertains different possible models of correlations will tend to behave more cautiously. An increase in the set of possible correlation structure may lead to more cautious behaviour.

To illustrate the implication of the tension between these two effects we analyse several applications. We consider a simple investment application in which investors observe past investments by others and reassess their positions. We show that when the number of investors is low, investors with a high correlation bound will reduce their initial risky positions due to the cautiousness effect described above. However, when the number of investors is large, the observation of others' beliefs might imply that the NB belief becomes more precise. In turn this will result in investors substantially increasing their risky behaviour. These results can shed light on behaviour before and after the 2008 financial crisis. First, the bounds on the correlation scenarios 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; a low volume of trade will indicate a less precise belief, resulting in more cautiousness. In addition, as mentioned above, CDO pricing might have changed drastically before and after 2008 due to similar forces.

Finally, we study the implications of our model to group decisions. We study a jury of individuals that deliberate (that is, exchange their beliefs) and then vote. This is indeed an environment in which individuals are exposed to the same evidence and hence the perception of correlation is relevant. Moreover, jurors are obliged to deliberate and exchange information. We show how the decisive voter is determined by their correlation bounds as well as by their preferences (threshold of doubt). We contrast the normative properties of our model to those of the literature. We show that juries can both over-acquit as well as over-convict and that jury size affects decisions in novel ways.

Our results contribute to several strands of the literature. First, our results are complementary to Epstein and Halevy (2019) who also study the relation between uncertainty about correlation and ambiguity. They consider preferences over lotteries that could either depend on outcomes of draws from one urn or from two urns. They show in their experiments that subjects exhibit stronger aversion to ambiguity when considering lotteries over

Snowberg (2015) and Glaeser and Sunstein (2009). In contrast, Sobel (2014) shows that correlation neglect is not a necessary condition for extreme beliefs.

the two urns rather than over each one separately, implying ambiguity aversion with relation to correlation across urns. They consider a model which allows for multiple "sources" of uncertainty, capturing risk, bias and correlation. Our model, in the prism of their analysis, considers predictions of decision makers who already observe draws from different individual "urns", and therefore focuses only on ambiguity over correlation.

Second, we contribute to a recent large literature on correlation neglect, i.e., a behavioral assumption that individuals neglect taking account of possible correlations between multiple sources of information.[4] 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. The papers on correlation neglect assume, in some environments, a misspecified model held by the decision makers, inducing wrong beliefs. In our framework the decision maker has a set of potentially misspecified models, with bounded degrees of correlation. Our results show that not only this too can end with a wrong belief, but that the wrong belief is the one associated with complete correlation neglect. That is, even when decision makers consider bounded correlation, when the naïve interpretation of the data is very informative, they all behave as if they have correlation neglect.

Finally, our results contribute to the literature in finance, which has since the 2008 crisis, considered extensively the issue of the uncertainty about correlation in default rates as well as across stress tests.[5] Our framework rationalises the procedures employed by rating agencies and investment banks when these evaluate complex assets. Risk analysis in financial firms that evaluate CDOs uses the individual level data of the loans making up a CDO, and then considers different correlation scenarios. As the FCIC report documents, "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." Indeed in practice, the set of scenarios that

[4]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. Alternatively, Ellis and Piccione (2017) use an axiomatic approach to represent decision makers affected by the complexity of correlations among the consequences of feasible actions. More generally, there is a recent literature on misspecified models, see for example Esponda and Pouzo (2016), Bohren (2016) and Heidheus et al (2018).

[5]Duffie et al (2009), Brunnermeier (2009), Coval et al. (2009), and Ellis and Piccione (2017), examine the effects of such misperceptions on financial markets.

are considered by this analysis has a particular structure; treating forecasts as independent is often used as a benchmark.[6] Around this benchmark, the level of correlation implicit in these scenarios is typically bounded. Tractability and simplicity imply that the different models used for generating correlation only allow for modest levels of correlation, using a small number of correlation parameters. When dealing with large numbers of components (e.g., the number of loans in a CDO), this implies bunching many assets to have the same correlation patterns (the "homogenous pool" problem).

Our results indicate that cautious or risky financial decisions can be a result of the tension between ambiguity aversion and the Naïve-Bayes updating rule. There are many anecdotes illustrating that decision making in organizations is sometimes akin to ambiguity attitudes, where either "optimists" or "pessimists" prevail. Anil K Kashyap's paper prepared for the FCIC observes that before the crisis there was an: "...inherent tendency for the optimists about the products to push aside the more cautious within the organization". After the crisis became apparent, "pessimism" prevailed: "Moody's officials told the FCIC they recognized that stress scenarios were not sufficiently severe...analysts took the "single worst case" from the M3 Subprime model simulations and multiplied it by a factor in order to add deterioration."[7]

2 The model

In this Section we present a theoretical model in which we define: (i) what a decision maker observes, namely the set of forecasts; (ii) how she uses a joint information structure to rationalise a set of forecasts and form a prediction on the state of the world; (iii) the level of ambiguity she faces over joint information structures. We then use this model to derive

[6] The Naïve-Bayes classifier, a method to analyse data by assuming different aspects of it are independent, is one of the work horses of operations research and machine learning. Querubin and Dell (2017) document how this approach was employed by the US military in the Vietnam war to assess which hamlets should be bombed based on multidimensional data collected from each hamlet. For more on the Naive Bayes approach see Russell and Norvig (2003) and Domingo and Pazzani (1996).

[7] Related to what we do in this paper a few recent papers have assumed ambiguity over correlation in different applications. Jiang and Tian (2016) analyze a financial market in which investors have ambiguity about the correlation between assets. They derive results relating to the volume of trade and asset prices. Easley and O'hara (2009) look more generally at the role of ambiguity in financial markets.

the set of predictions a decision maker can reach when combining forecasts.

2.1 Information

We first describe the information of the decision maker, which includes the state space, some prior knowledge, and observed forecasts.

The decision maker knows the following aspects of the environment:

1. The state space. The state is a $n - tuple$ vector $\boldsymbol{\omega} =\{\omega_1, ..., \omega_n\}$, $\omega_i \in \Omega$, $\boldsymbol{\omega} \in \Omega^n$, where $n \geq 1$ and Ω is a finite space (in the Appendix we consider the case of continuous distributions which may be more suitable for some applications).

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 observes K forecasts. Specifically, there are k_i forecasts about each ω_i, so that $\sum_{i=1}^{n} k_i \equiv K$. A typical forecast j on element i is a (full support) probability distribution, $q_i^j(\omega_i)$, over Ω. Let $\mathbf{q}$ denote the vector of the K observable forecasts.

The model allows us to consider both correlations between forecasts (even when $n = 1$) as well as correlations across the different elements of the state (when $n > 1$). For example, correlation across forecasts arises or when banks that conduct stress tests persistently ignore the same type of information (or, in another application, when political pollsters' strategies systematically neglect parts of the population across US states). Correlation across the elements of the state arises for example when the returns of assets are correlated (or when the voting outcome across US states depends on a common shock).

The decision maker will combine these forecasts to reach a set of predictions about the state. A prediction about the state is a probability distribution η over Ω^n and we are interested in the set of rationalisable predictions, as 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, that is, a joint information structure.

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

1. A joint prior distribution, $p(\boldsymbol{\omega})$, for which the marginal on element i is $p_i(\omega_i)$.

2. A set of $K-tuple$ vectors of signals $S = \times_{i=1}^{n} \times_{j=1}^{k_i} S_i^j$, 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, $\hat{f}(\mathbf{s}, \boldsymbol{\omega})$, where $\mathbf{s} \in S$. Specifically, let $\hat{f}(\mathbf{s}, \boldsymbol{\omega}) = p(\boldsymbol{\omega})\hat{f}(\mathbf{s}|\boldsymbol{\omega})$, where $\hat{f}(\mathbf{s}|\boldsymbol{\omega})$ is the distribution over signals generated by $\boldsymbol{\omega}$. Also let $\hat{f}_i^j(s|\omega_i)$ denote the marginal information structure for source j on element i that is derived from $\hat{f}(\mathbf{s}|\boldsymbol{\omega})$.

Note that in a joint information structure, both the elements of the state can be correlated, through $p(\boldsymbol{\omega})$, and the signals generating the different forecasts could be correlated, through $\hat{f}(\mathbf{s}|\boldsymbol{\omega})$.

We are now ready to define formally a rationalisable prediction:

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

In other words, the decision maker can ratioanlize a set of forecasts by constructing a joint information structure and a set of signals so that each forecast can be derived by Bayes rule given the forecaster's signal, the prior over his assigned dimension of the state and the marginal distribution generating the signal. Using this set of signals and the joint information structure she can then generate a prediction on the state of the world. We will be interested in the set of predictions that can be rationalized given $\mathbf{q}$, and the set of joint information structures considered by the decision maker.

In the main part of the analysis we assume that the agent only observes the forecasts $\mathbf{q}$. In the Appendix we show that our results are robust to the agent also observing the signals and the marginal information structures of the different sources. Intuitively, the information gleaned from marginals and signals is still not sufficient to recover the structure of correlation, which is the main focus of our analysis.

We now define the set of joint information structures over which the decision maker faces ambiguity.

2.3 Ambiguity over correlation

We now provide a general and simple one-parameter characterization for a set of joint information structures with bounded correlation, which will define the level of ambiguity the decision maker faces. To this end, we use the exponent of the *pointwise mutual information* (ePMI) to define bounds on the correlation between information structures. Specifically, throughout the paper, we assume the following:

Assumption A1: *There is a parameter* $1 \leq a < \infty$, *such that the decision maker only considers joint information structures,* $(S,\ \Omega^n, p(\boldsymbol{\omega}),\ \hat{f}(\mathbf{s}, \boldsymbol{\omega}))$, *so that at any state* $\boldsymbol{\omega} \in \Omega^n$ *and for any vector of signals* $\mathbf{s} \in S$,

$$\frac{1}{a} \leq \frac{\hat{f}(\mathbf{s}, \boldsymbol{\omega})}{\prod_{i=1}^{n} p_i(\omega_i) \prod_{j=1}^{k_i} \hat{f}_i^j(s_i^j|\omega_i)} \leq a.$$[8]

The ratio $\frac{\hat{f}(\mathbf{s}, \boldsymbol{\omega})}{\prod_{i=1}^{n} p_i(\omega_i) \prod_{j=1}^{k_i} \hat{f}_i^j(s_i^j|\omega_i)}$ is the (exponent) of the pointwise mutual information; this is the ratio of the joint probability of some $\mathbf{s}$ and $\boldsymbol{\omega}$ to the probability generated by its marginals when we assume independence. If $a = 1$, then at any point $\mathbf{s}$ and $\boldsymbol{\omega}$, this ratio equals 1, implying that the decision maker considers only joint information structures that satisfy independence. When a is larger than one, this implies that the decision maker considers some correlation or in other words, that there is mutual information across the variables at some $\mathbf{s}$ and $\boldsymbol{\omega}$ (see also Example 1 below). The parameter a, the ePMI-bound, describes therefore the extent of the ambiguity the decision maker faces over the set of correlation scenarios. The larger is a, the larger is the set of joint information structures that satisfy A1. Thus, ambiguity is larger when a is larger.

The formulation of the set is general, detail-free in terms of the underlying distribution functions, and captures the maximal set of joint information structures with correlation bounded by a. Note also that in different environments, individuals or organizations may be able to have different such sets (for example, a may depend on the number of sources K).

It is often the case that the sets over which decision makers have ambiguity contain the truth; our model will be general in the sense that pitted against the rational decision maker who is aware of the true joint information structure, the decision maker may consider less correlation or more correlation. We discuss this in our applications.

[8]All the results can be easily generalized if instead of the lower bound $\frac{1}{a}$ we use some finite $b < 1$.

The example below illustrates the relation between ePMI, the level of a, and correlation. Specifically, we show how the ePMI values of a joint information structure must take values which are both below and above 1, and hence the set of such joint information structures will always include the case of (conditional) independence.

Example 1: Assume $\Omega = \{0, 1\}$, two states, ω_1 and ω_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 $\hat{f}(s|\omega)$ satisfies independence, but that the prior satisfies correlation, as described in the following symmetric matrix, where $p \equiv p(0)$:

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

When $\varepsilon = 0$, the ePMI equals 1 at any state. If ε is positive then we have positive correlation across the states, whereas if ε is negative, we have negative correlation.

Suppose for the sake of exposition, that the decision maker only considers a positive ε. Note now that the ePMI at $\boldsymbol{\omega} = (0, 0)$ is $\frac{p^2+\varepsilon}{p^2} > 1$, whereas the ePMI at $\boldsymbol{\omega} = (0, 1)$ is $\frac{p(1-p)-\varepsilon}{p(1-p)} < 1$. This is a general property: 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, to maintain this as a distribution function. Thus fixing the ePMI at 1 is in some sense the simplest possibility.

Moreover, note that the ePMI constraints for a positive ε,

$$\frac{p(1-p) - \varepsilon}{p(1-p)} \geq \frac{1}{2}, \ \frac{p^2 + \varepsilon}{p^2} \leq a, \ \frac{(1-p)^2 + \varepsilon}{(1-p)^2} \leq a,$$

imply, assuming wlog that $p > \frac{1}{2}$, and that $a \geq \frac{p}{1-p}$, that:

$$\varepsilon \leq p(1-p)(1 - \frac{1}{a}).$$

Note that for the above example, the correlation coefficient between the two states is $\rho_\varepsilon = \frac{\varepsilon}{p(1-p)}$. As we need $\varepsilon \leq p(1-p)$ for the above to be a joint distribution function, we have that the correlation coefficient satisfies $\rho_\varepsilon \leq 1$. Given the ePMI constraints however, the decision maker considers only $\rho_\varepsilon \leq 1 - \frac{1}{a}$, and hence bounded correlation.

Our formalisation in Assumption A1 implies that the ambiguity set contains a large set of information structures around $\rho = 0$ (therefore including also negative correlation in the example above). In some applications,see Jiang and Tian (2016), one can focus on sets of ambiguity beliefs around a particular ρ. Some of our results below, such as Observation 1, may apply also to these cases as long as the set considered falls within those that satisfy assumption A1.

Let $C(a, \mathbf{q})$ be the set of beliefs $\eta(.)$ that are rationalisable, as in Definition 1, given the vector of forecasts $\mathbf{q}$, by information structures that satisfy A1 for some ePMI-bound a. In other words, the decision maker considers each joint information structure in her set, one by one, and for each derives a rationalisable prediction (if feasible). Our main result below characterizes $C(a, \mathbf{q})$.

In the remainder of this Section we explain how the ePMI captures correlation. In short, the average of the PMI is the well known *mutual information* measure, and moreover, it implies bounded *concordance*, which is the most general non-parametric measure of correlation. It therefore allows for a more general relation between variables than is typically captured by assumptions such as linear correlation.

2.4 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 $g(x_1, ..., x_n)$ be a joint probability distribution of random variables $\tilde{x}_1, ..., \tilde{x}_n$, with marginal distributions $g_i(.)$. The pointwise mutual information (PMI) at $(x_1, ..., x_n)$ is $\ln[\frac{g(x_1,...,x_n)}{\Pi_i g_i(.)}]$. For example, for two variables, it can also be written as

$$\ln[\frac{g(x_1, x_2)}{\Pi_i g_i(.)}] = h(x_1) - h(x_1|x_2)$$

where $h(x_1) = -\log_2 \Pr(\tilde{x}_1 = x_1)$ 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, $MI(X_1, X_2) = \sum_{x_1 \in X_1} \sum_{x_2 \in X_2} g(x_1, x_2) \ln[\frac{g(x_1,x_2)}{\Pi_i g_i(.)}] = H(X_1) - H(X_1|X_2)$, which is 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:

$$MI(X_1, X_2) = D_{KL}(g(x_1, x_2)|g_1(x_1)g_2(x_2)),$$

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.[9]

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. Concordance measures how well the relationship between two variables can be described using a monotonic function.[10] In the Appendix we show (Proposition B1) how a bounded PMI translates to a bounded concordance measure.

3 The Main Result: Characterising $C(a, \mathbf{q})$

In this Section we characterise $C(a, \mathbf{q})$, the set of rationalisable beliefs of the decision maker that are derived from the set of joint information structure she considers, as defined in A1, for an ePMI-bound a. Specifically, we are interested in understanding how ambiguity over information sources translates into ambiguity over the state of the world.

Let $\eta^{NB}(.)$ denote a posterior belief of an individual who uses a Naive-Bayes approach, that is, she believes that there is no correlation across the states or the information sources. Our result below implies that this NB belief is a useful tool for the characterisation of the whole set of belief, as Proposition 1 establishes.

Proposition 1: *Suppose that $n > 1$ or $K > 1$. A belief $\eta(.)$ is in $C(a, \mathbf{q})$ for $1 \leq a < \infty$ if and only if it satisfies*

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

[9] The PMI therefore does not distinguish between rare or frequent events.

[10] The most common measure of concordance is Spearman's rank correlation coefficient. A perfect Spearman correlation of +1 or -1 occurs when each of the variables is a perfect monotonic function of the other.

for a vector $\boldsymbol{\lambda} = (\lambda_{\boldsymbol{\omega}})_{\boldsymbol{\omega} \in \Omega^n}$ *satisfying* $\lambda_{\boldsymbol{\omega}} \in [\frac{1}{a}, a]$ *for all* $\boldsymbol{\omega}$, *where*

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

Moreover, the set $C(a, \mathbf{q})$ *is compact and convex.*

The result shows that there are two sufficient statistics that allow us to characterize the set of beliefs held by the agent: the PMI-bound a and the NB posterior $\eta^{NB}(.)$. Thus, while the decision maker is faced with a complicated environment, her Bayesian combined forecasts can be derived with a simple heuristic-like behavior. She needs to consider the Naïve-Bayes benchmark, as if she neglects correlation, and to 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. We discuss in Section 4.2 how a modeler can also identify a.

As can be seen from the characterisation, when $a = 1$ then $C(1, q)$ is a singleton and contains a unique belief $\eta^{NB}(.)$. However, when $a > 1$, the set of beliefs is not unique once we have multiple forecasts. When the decision maker considers also joint information structures with some level of correlation, she can rationalise a larger set of beliefs about the state of the world. Thus, ambiguity over joint information structures now translates into ambiguity over the state of the world. This arises only when the decision maker considers correlation, so that $a > 1$, and when there is more than one forecast. Specifically, when there is more than one forecast, then the different levels of correlation considered "kick" in to induce different beliefs, while when only one forecast exists, this effect does not arise (in that case, by rationalisability, her prediction is the unique forecast she is exposed to). Thus, the decision maker becomes less confident in terms of facing larger ambiguity over the state of the world when she considers correlation and when she has more than one forecast.[11]

[11] This is related to the notion of dilation introduced in Seindenfeld and Wasserman (1993). Seindenfeld 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 she gains after observing multiple sources, sometimes dilation occurs. See also Bose and Renou (2014) and Epstein and Schneider (2007).

3.1 Sketch of the proof

Necessity: It is easy to see the necessary part of the proof. Specifically, for any $\mathbf{s}$ and a joint information structure, the ePMI constraints imply:

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

By rationalisability however:

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

As a result:

$$\frac{\frac{1}{a}\prod_{i=1}^{n}\frac{\prod_{j=1}^{k_i}q_i^j(\omega_i)}{p_i(\omega_i)^{k_i-1}}}{a\prod_{i=1}^{n}\frac{\prod_{j=1}^{k_i}q_i^j(\omega_i')}{p_i(\omega_i')^{k_i-1}}} \leq \frac{p(\boldsymbol{\omega})\hat{f}(\mathbf{s}|\boldsymbol{\omega})}{p(\boldsymbol{\omega}')\hat{f}(\mathbf{s}|\boldsymbol{\omega}')} \leq \frac{a\prod_{i=1}^{n}\frac{\prod_{j=1}^{k_i}q_i^j(\omega_i)}{p_i(\omega_i)^{k_i-1}}}{\frac{1}{a}\prod_{i=1}^{n}\frac{\prod_{j=1}^{k_i}q_i^j(\omega_i')}{p_i(\omega_i')^{k_i-1}}}.$$

Note that when $a = 1$, there is then a unique feasible belief, $\eta^{NB}(.)$.

Sufficiency: The proof in the Appendix shows the sufficiency of the characterisation. We show sufficiency by constructing an information structure that yields each belief in the set and satisfies the rationalisability and ePMI constraints. Specifically, for any vector $(\lambda_\omega)_{\omega\in\Omega}$ that satisfies $\frac{1}{a} \leq \lambda_\omega \leq a$ for any realisation of ω, we construct an information structure that induces the belief (suppose that $n = 1$ for simplicity but $K > 1$) :

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

To do this, we define the probability that each information source j receives a signal s^* in state $\omega \in \Omega$, as $\delta_\omega^j = \varepsilon\frac{q^j(\omega)}{p(\omega)}$ for some $\varepsilon > 0$, and define the probability that all sources receive s^* in state $\omega \in \Omega$ as $\alpha_\omega = \lambda_\omega\prod_{j\in K}\delta_\omega^j$. It is easy to see from this that indeed

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

and moreover that

$$\frac{p(\omega)\delta_\omega^j}{\sum\limits_{\omega'\in\Omega} p(v)\delta_{\omega'}^j} = \frac{q^j(\omega)}{\sum\limits_{\omega'\in\Omega} q^j(\omega')} = q^j(\omega),$$

which implies that the posterior beliefs of all individuals are rationalized. It is also easy to see that at any ω, when all receive s^*, the ePMI constraint is $\frac{\alpha_\omega}{\prod\limits_{j\in K}\delta_\omega^j} = \lambda_\omega$ and is hence satisfied. By using small enough values of ε, we are able to construct an information structure that satisfies the ePMI constraints at all other points, that is, for any $(\omega, s^1, ..., s^K)$, where $s^j \in \{s^*, s^{-*}\}$.

Convexity of $C(a, \mathbf{q})$: It is not straightforward to show convexity of the joint information structures; due to the nature of the ePMI constraints, one cannot simply take a convex combination of joint information structures where each rationalises a belief in $C(a, \mathbf{q})$ in order to rationalize a convex combination of beliefs. To prove convexity we therefore use our characterization of the set of beliefs beliefs in $C(a, \mathbf{q})$. Specifically, consider two beliefs η and η' in $C(a, \mathbf{q})$. Then for any $\beta \in [0, 1]$, we know that $\frac{\beta\eta(\boldsymbol{\omega})+(1-\beta)\eta'(\boldsymbol{\omega})}{\beta\eta(\boldsymbol{\omega}')+(1-\beta)\eta'(\boldsymbol{\omega}')} \leq \frac{\beta\eta(\boldsymbol{\omega}')+(1-\beta)\eta'(\boldsymbol{\omega}')}{\beta\eta(\boldsymbol{\omega}')+(1-\beta)\eta'(\boldsymbol{\omega}')} a^2 \frac{\eta^{NB}(\boldsymbol{\omega})}{\eta^{NB}(\boldsymbol{\omega}')} = a^2 \frac{\eta^{NB}(\boldsymbol{\omega})}{\eta^{NB}(\boldsymbol{\omega}')}$, where we use $\frac{\eta(\boldsymbol{\omega})}{\eta(\boldsymbol{\omega}')} \leq a^2 \frac{\eta^{NB}(\boldsymbol{\omega})}{\eta^{NB}(\boldsymbol{\omega}')}$, and $\frac{\eta'(\boldsymbol{\omega})}{\eta'(\boldsymbol{\omega}')} \leq a^2 \frac{\eta^{NB}(\boldsymbol{\omega})}{\eta^{NB}(\boldsymbol{\omega}')}$, as both beliefs are in $C(a, \mathbf{q})$. The lower bound is similarly attained. Thus, convexity is proved directly on the set of beliefs.

3.2 The Naïve-Bayes and cautiousness effects

The characterisation of the maximal set of beliefs allows us to make two simple observations. The first observation -which we call the Naïve-Bayes effect- is that if the NB belief is very precise, then the set $C(a, \mathbf{q})$ will in some cases coincide with it, implying that individual will behave as if she has correlation neglect. The second observation is that $C(a, \mathbf{q})$ is larger when a is larger. In the presence of ambiguity aversion, this may imply greater cautiousness. In the next Section we show how the interaction between these two effects induces sometimes risky and sometimes cautious shifts in investment behaviour.

The Naïve-Bayes effect: The characterisation in Proposition 1 allows us to see how the precision of $\eta^{NB}(.)$ affects the size of $C(a, \mathbf{q})$. Consider the case where $\eta^{NB}(.)$ is very precise (but not necessarily correct). This could arise for example when the number of forecasts K grows large and when the NB belief converges to be degenerate. Consider a sequence of decision making problems with a sequence of vectors of forecasts $\mathbf{q}_K$ and a sequence of ambiguity sets characterised by a_K. These imply a sequence of NB beliefs $\eta^{NB}_{\mathbf{q}_K}(.)$ and sets of beliefs $C(a_K, \mathbf{q}_K)$ which are the focus of the observation below.

Observation 1: *Suppose that there exists a $\boldsymbol{\omega}' \in \Omega^n$ such that $\eta^{NB}_{\mathbf{q}_K}(\boldsymbol{\omega}') \rightarrow_{K\rightarrow\infty} 1$. If $\lim_{K\rightarrow\infty}(a_K)^2(1-\eta^{NB}_{\mathbf{q}_K}(\boldsymbol{\omega}')) = 0$ then $C(a_K, \mathbf{q}_K)$ converges to the singleton belief which is the degenerate belief on $\boldsymbol{\omega}'$.*

To see this, note that by Proposition 1 we have that with n states and K information sources, $\eta(.) \in C(a_K, \mathbf{q}_K)$ is rationalisable if and only if it satisfies

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

for a vector $\lambda = (\lambda_{\boldsymbol{\omega}})_{\boldsymbol{\omega}\in\Omega^n}$ satisfying $\lambda_{\boldsymbol{\omega}} \in [\frac{1}{a_K}, a_K]$ for all $\boldsymbol{\omega} \neq \boldsymbol{\omega}'$. Note that $\frac{\lambda_{\boldsymbol{\omega}}}{\lambda_{\boldsymbol{\omega}'}} < (a_K)^2 \frac{1-\eta^{NB}_{\mathbf{q}_K}(\boldsymbol{\omega}')}{\eta^{NB}_{\mathbf{q}_K}(\boldsymbol{\omega}')}$ but as $\lim_{K\rightarrow\infty}(a_K)^2(1-\eta^{NB}_{\mathbf{q}_K}(\boldsymbol{\omega}')) = 0$ this implies that $\frac{\eta(\boldsymbol{\omega})}{\eta(\boldsymbol{\omega}')}$ has to converge to zero.

The observation above illustrates that even when the decision maker considers information structures with high degrees of correlation, a very precise NB belief may overwhelm considerations of correlation. As our set of predictions is the maximal set for all rational decision makers who consider bounded correlation, the observation that the set of predictions can shrink to a singleton NB belief implies that behaviour à la correlation neglect can arise for all types of assumptions on the decision maker. In other words, even when the decision maker does not have ambiguity over the set of joint information structures, but has a prior over these, the result is the same. Or, alternatively, if the decision maker considers only a subset of correlation structures which does not include independence, the limit result would be that her beliefs would still converge to the NB belief, as the whole set of predictions -which would include her predictions for any K, will shrink to that belief. The following example illustrates the fact that the decision maker can satisfy the condition in the observation above while still considering, ex ante, high degrees of correlation.

Example 2: In this example we provide an information structure that satisfies: (i) $a_K \rightarrow \infty$, along with large degrees of ex ante correlation across signals for all K;[12] (ii)

[12] Note that a_K captures ex post correlation as it is the bound on the ePMI at every point, that is, for any states and forecasts.

$\lim_{K\to\infty}(a_K)^2(1-\eta^{NB}_{\mathbf{q}_K}(\boldsymbol{\omega}')) = 0$. Specifically, consider just one state, $\omega \in \{0,1\}$, with a uniform prior. The decision maker believes she receives predictions according to the following information structure: With probability $\alpha > 0$, the signals are correlated as explained below, and with probability $1-\alpha$ each signal j is drawn independently, where $s^j \in \{0,1\}$ and $\Pr(s^j = \omega|\omega) = q > \frac{1}{2}$. In the correlated event, when $\omega = 1$, with probability α, a number qK of the signals are randomly chosen and are assigned to have $s^j = 1$ and the remaining are assigned with $s^j = 0$ (choose q so that qK is a number). When $\omega = 0$, with probability α, a number qK is drawn to assigned to have $s^j = 0$ and the remainder is assigned to have $s^j = 1$. Thus the marginal probability of each signal to provide the correct realisation is q. Note that with probability α the realisations of the signals are highly correlated (and provide information that is equivalent to that of one signal only). Specifically, the correlated event is chosen to mimic the most likely events under independence. Intuitively then, the decision maker here is "suspicious" at distributions of realisations that mimic the most likely events under independence, and believes that these arise from correlation. We show in the Appendix that for this information structure, $a_K = \alpha\sqrt{2\pi}\sqrt{Kq(1-q)} + (1-\alpha) \to_{K\to\infty} \infty$, while $\lim_{K\to\infty} a_K^2(1-\eta^{NB}_{\mathbf{q}_K}(1)) = 0$, where $\mathbf{q}_K$ is the vector of forecasts in which all entries are q. Specifically, this arises as a_K becomes large in the slowest rate possible,[13] while $\eta^{NB}(.)$ becomes degenerate in the fastest rate possible. The example illustrates that the condition derived in Observation 1 builds on the fact that the upper bound for a_K might be achieved at a different event to the event that the individual actually observes in the (ex-post) data.[14]

Note that the NB benchmark -while relying on many pieces of information- can still differ substantially from the rational belief given the true joint information structure. For example, the true probability distribution could be that the forecasts are positively correlated -either all q or all $1-q$- with a probability close to one. A rational prediction upon observing a vector of forecasts in which all entries are q, as considered in Example 2, implies then only a belief q that the state is 1, while in the Example above the belief would be the degenerate NB belief. Moreover, this true information structure would generate the observation described in Example 2 (a vector of forecasts in which all entries are q) with a high likelihood.

[13] This is the case as the highest degree of correlation arises for an event which is the most likely under independence.

[14] The example can easily be generalised to cases in which upon observing the ex-post data the individual always entertains some level of correlation, but in the limit converges to believe the data is independent.

Put differently, note that a decision maker who considers the true information structure will not arrive at a wrong prediction. Thus, when the true correlation structure has bounded levels of correlation in the limit, then indeed it cannot be that $\eta^{NB}_{\mathbf{q}_K}(\boldsymbol{\omega}')$ will become degenerate on the wrong state (as there would be sufficient independent information to enable correct NB-belief). However, when this is not the case, then a decision maker who is aware of the true information structure will need to consider a_K so that $\lim_{K\to\infty}(a_K)^2(1-\eta^{NB}_{\mathbf{q}_K}(\boldsymbol{\omega}')) > 0$.

The cautiousness effect: The Proposition also unravels a simple relation between confidence and correlation:

Observation 2: *If* $a < a'$, *then* $C(a, \mathbf{q}) \subset C(a', \mathbf{q})$.

In our model individuals who consider a larger set of joint information structures enabling for greater degrees of correlation, will end up with a larger set of predictions. Thus, considering more joint information structures will reduce confidence in the sense that individuals may not be sure what is the right belief. Fixing their level of ambiguity aversion, or attitudes to ambiguity, considering an increased level of correlation can induce a more cautious behaviour, as we show below. Specifically, for any $\eta^{NB}(.)$, 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 can explain pessimistic behavior in financial markets when investors believe they face unknown levels of correlation as we now explore.

4 Applications

In this Section we consider three applications. We first illustrate how the NB effect influences CDO rating, especially when the CDO is complex. We then consider an investment environment where investors can observe others' behaviour. In this application we highlight the interaction between the volume of trade and investor confidence that results from the NB and the cautiousness effects. Finally, we consider an application to a group decision making (juries) and show how group decisions can further be distorted when jurors are uncertain about the correlation in the evidence they observe.

4.1 Complex CDO rating

In this Section we provide a simple model of risk management for CDOs. Relying on Observation 1, we will show that for complex CDOs, for any ePMI-bounded dependence structure across loans that one considers, the CDO can receive the highest rating.

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 α of the individual loans default.

In this application the uncertainty over correlation will be about the correlation between the defaults of the individual loans, i.e., through the prior p over the state. 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 equal to 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 $\lim_{n\to\infty} \frac{\sum_{i=1}^{n} p_i}{n} = \mu < \infty$. Again, we assume a to be fixed although the result can be extended to consider a sequence a_n.

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.[15] We instead describe ePMI bounds without resorting to any functional forms.

By Proposition 1, for any state $\boldsymbol{\omega}$, we have that a belief $\eta(.)$ is in the set $C(a, \mathbf{q})$ iff:

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

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

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

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

Let us now consider the worst case scenario among the scenarios determined by the extent of correlation a. Using Proposition 1 we can derive the following:

[15] See for example Wang et al (2009).

Remark 1: *The worst-case scenario is that the CDO fails with probability*

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

As a CDO fails in many environments, Remark 1 allows us to see that our analysis easily carries through for *a combination of states*. We can now use this to formalise CDO rating.

We focus on what rating is awarded when the CDO is complex, i.e., when n is large. To compute the probability of default under no correlation for a large n, we can approximate the Poisson Binomial distribution with a Poisson distribution with a mean μ.[16] This implies that in the limit, under independence, there will be a share $\lceil \mu n \rceil$ of assets that would fail. As a result, a rating agency which considers no correlation, will, for a large n, approximate $\sum_{l=\lceil \alpha n \rceil}^{n} \sum_{\boldsymbol{\omega}^l \in \Omega^l} \eta^{NB}(\boldsymbol{\omega}^l)$, the probability that the CDO fails, by 0 if $\alpha > \mu$, and by one if $\alpha < \mu$. In other words, while the probability of each feasible state $\boldsymbol{\omega}$ does not become degenerate, 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.

Suppose that the rating agency chooses a triple A rating to the CDO if its probability of default is lower than some exogenous cutoff $x > 0$. Given the above discussion, we can then establish:

Proposition 2: For any $x > 0$ and $a < \infty$, *if* $\alpha > \mu$, *there is a large enough* n *such that the CDO receives the highest rating.*

Given that the NB-belief converges to be degenerate for large n, the extent of correlation bound a is immaterial, and even rating agencies that consider the worst case scenario will, when $\alpha > \mu$, award the highest rating for any level of x. In other words, neither a nor x are important for the rating rule.

This implies that with complex securities composed of many assets, there are environments in which taking bounded correlation into consideration will not change investors' behaviour. Even if the pessimists get their say in an organization, their recommendation would be to provide a high rating. We therefore unravel a relation between complexity and correlation neglect. Note that in financial markets, it is common to consider bounded levels of correlation. Tractability and simplicity imply that the different models used for generating

[16] See Hodges and Le Cam (1960).

correlation only allow for modest levels of correlation, using a small number of correlation parameters. When dealing with large numbers of components (e.g., the number of loans in a CDO), this implies bunching many assets to have the same correlation patterns.[17]

To illustrate the implication of Proposition 2, let us consider the following structure that generates the true correlations between loans. Suppose that there is a default parameter p, taken from some $f(p)$ on $[0,1]$, with $E_f(p) = \bar{p}$. When p is drawn, then a $\lceil pn \rceil$ of the assets fail. The rating agency knows the true marginal probability of each asset to fail, which is $p_i = \bar{p}$, and considers different correlation scenarios around this marginal. We then have:

Proposition 3: *For large n, CDOs receive the wrong triple A rating whenever* $1-F(\alpha) > x$ and $\bar{p} < \alpha$, and the wrong lower rating whenever $1 - F(\alpha) < x$ and $\bar{p} > \alpha$.

To see why this arises, note that the rating agency, as in Proposition 2, will award the highest rating when $\alpha < \mu$, and the lowest when $\alpha > \mu$. Given that the marginal probability of each asset to fail is $\bar{p}$, and thus the average failure probability μ for the Poisson approximation is $\bar{p}$, a triple A rating is awarded if and only if $\bar{p} < \alpha$. However, given the true information structure which exhibits full correlation, the probability that at least a share of α assets fail is simply the probability that $p \geq \alpha$, that is, $1 - F(\alpha)$. This is because whenever p is drawn, a share $\lceil pn \rceil$ defaults, and hence when $p \geq \alpha$ the CDO will fail. As a result, efficient rating implies a triple A rating, according to the exogenous cutoff x, if and only if $1 - F(\alpha) < x$. Thus, a triple A rating is wrongly awarded when both $\bar{p} < \alpha$ and $1 - F(\alpha) > x$, and is wrongly avoided when $1 - F(\alpha) < x$ and $\bar{p} > \alpha$. Suppose for example that f is uniform. Then efficient triple A rating should be awarded when $1 - \alpha < x$ so that $\alpha > 1 - x$, while instead the CDO is awarded a triple A rating whenever $\alpha > 0.5$.

4.2 Risky and cautious investment shifts

We now consider a simple investment model to highlight the effect of the interaction of Observations 1 and 2. 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

[17] Similarly, the families of correlation structures (copulas) that are often used, implicitly limit the levels of correlation. Also, arbitrary historical correlation data, which typically exhibits moderate levels of correlation over time, is often used to generate scenarios. See MacKenzie and Spears (2014) and the FCIC report mentioned above.

state, and a risky asset which provides 0 at state 0 and $H > L$ in state 1. Each investor has one unit of income to invest which she can split across these two assets. Assume a standard concave utility $V(.)$ 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.[18]

There are k informed investors. Let each hold a prediction $q^j(\omega)$. Thus, in the first period, they invest according to $q^j(1)$.[19] Investments in the first period are observed; assume that investors can then backtrack the beliefs of others, $q^j(1)$ for all j.[20] Finally, in the second period, the investors can adjust their investments following their observation of $\mathbf{q}$.

For simplicity we assume that the PMI-bound of each investor j, a^j, is fixed. As Observation 1 shows, our results will remain if we consider that these depend on k, as long as they do not converge too fast to infinity. To take into consideration the cautiousness effect that can arise with ambiguity as described following Observation 2, we assume here that when individuals are faced with ambiguity, they use the max-min preferences as in Gilboa and Schmeidler (1989).[21] Thus, in the second period, after observing others' investments, an individual j with ambiguity aversion will then base her investment decision on the belief which minimises her utility, which is $\min_{\eta^j(1)\in C(a^j,\mathbf{q})} \eta^j(1)$.

Given Proposition 1, we can further simplify $\eta^{NB}(.)$. 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., $\frac{\hat{q}(1)}{1-\hat{q}(1)} = (\Pi_{j\in K}\frac{q^j(1)}{1-q^j(1)})^{\frac{1}{k}}$. We can now express

[18] Here we abstract from prices. See the discussion at the end of the section.

[19] We can assume that each investor receives a signal s^j on ω, knows the marginal $q^j(s^j|\omega)$, and updates her prediction to $q^j(\omega)$.We show in Appendix B how all our analysis also holds when the agents who combine forecasts also have their own information. 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 Proposition 1, with the knowledge of the marginals and signals (see Appendix B).

[20] This assumption is made here for simplicity. One can 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.

[21] 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.

$\eta^{NB}(1)$ as (recall that k is the number of investors),

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

And $j's$ minimum combined forecast can be then written as:

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

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

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

Proposition 4: *In the second period, following exposure to* $\mathbf{q}$*: (i) If* $a^j < a^{j'}$*, then 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 her investment in the risky asset compared to her first period's investment; (iii) For a large enough* k*, if* $\hat{q}(1) > \frac{1}{2}$*, all investors will increase their investment in the risky asset compared to their first period's investment and if* $\hat{q}(1) < \frac{1}{2}$*, then all will decrease their investment in the risky asset.*

Part (i) illustrates that individuals who consider a smaller set of correlated information sources will behave in a more risky manner.[22] This result also implies that we can identify the individual PMI-bounds from choice data, as long as there is general data on behavior of individuals in the face of ambiguity. Specifically, assume that given some ambiguous set of priors over the state of the world, we isolate some individuals with max-min behaviour. That is, these individuals invest according to the most conservative prior. Now take these max-min individuals and present them with a unique prior and a set of forecasts. Different investment behaviours should reflect then different views of correlation; thus we can differentiate those who consider lower levels of correlation by their more risky behavior.

To see how parts (ii) and (iii) arise, recall that we have identified the cautiousness effect and the Naïve-Bayes effect. If beliefs are in general pessimistic (that is, $\hat{q}(1) < \frac{1}{2}$), then

[22] Note that "standard" results in the literature on 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 (see Glaeser and Sunstein 2009 or Ortoleva and Snowberg 2015). 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.

both go in the same direction inducing a cautious investment 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. When the number of forecasts is small, the Naïve-Bayes effect is weak, as $\eta^{NB}(.)$ is not likely to be informative. We can then always find a high enough level of ambiguity so that the cautiousness effect will dominate.[23] On the other hand, when the number of forecasts is large, $(\frac{\hat{q}(1)}{1-\hat{q}(1)})^k$ becomes very large, and the NB belief overcomes cautiousness to induce a substantial risky behaviour. To recap, confidence can arise when a is small or $\eta^{NB}(.)$ is relatively precise. On the other hand, cautiousness arises when a is large, and $\eta^{NB}(.)$ is relatively imprecise.

Remark 2: While cautiousness is directly related to ambiguity, a cautious shift will not always arise with "standard" forms of ambiguity. Suppose for example that individuals have ambiguity over the prior in the set $[\frac{1}{2} - \varepsilon^i, \frac{1}{2} + \varepsilon^i]$ for some ε^i. Suppose that the information structures satisfy independence and that this is known, and that all individuals start from some beliefs $q^i(1) = q > \frac{1}{2}$, as above. Following the first period, individuals will always become more optimistic and increase their level of investment. Ambiguity over the prior implies that first and second period investment are both lower compared to the case of no ambiguity, but that second period investment increases for any k.

Both of the effects we unravel can potentially shed light on the behaviour of investors before and after the 2008 financial crisis. Many investors had realized after 2008 that the level of correlation in assets and across forecasts was much higher than initially perceived. In response, as we document in the introduction, the worst case scenarios did not only receive more weight in the overall assessment, but were also downgraded to capture a more pessimistic outlook. This corresponds to a possible shift of the value of a which, as we show, can contribute to a "confidence crisis" and lower investment levels.

Another element that changes in the market is the informativeness of $\eta^{NB}(.)$ which depends on the number of investors involved. A market with many investors (even small ones) is such that individuals can observe many forecasts. Even if each investment is slightly optimistic, it can be aggregated to a precise and very optimistic $\eta^{NB}(.)$, which will overshadow the

[23] This result 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).

cautiousness effect. On the other hand, once some skepticism arises, as happened after the crisis, the market will consist of less investors. In this case, the benchmark $\eta^{NB}(.)$ would be imprecise, and the cautiousness effect will dominate.[24]

4.3 Jury decision making

In our final application we use our framework to study decision making in Juries. This is indeed an environment in which individuals are exposed to the same evidence and hence the perception of correlation is relevant. Moreover, jurors are obliged to deliberate and exchange forecasts.[25] The analysis in this application illustrates how observations 1 and 2 interact in a *collective* decision making problem.

We adopt the canonical model of juries to our framework. There are n jurors who decide to acquit, A, or convict, C, a defendant. There are two states of the world, one in which the defendant is guilty, G, and another in which she is innocent, I. Assume that the prior about these two states is uniform and that it is common knowledge.

As in our main model, a vector of beliefs, $\mathbf{q}(G) = (q^1(G), ..., q^n(G))$, denotes the posterior private belief of each juror that the defendant is guilty, attained after observing their signals from the trial. During deliberation the jurors share their beliefs $q^i \equiv q^i(G)$ truthfully. Each Juror has a correlation parameter $a^i > 1$, satisfies A1, and updates her beliefs to the set characterised in Proposition 1. Each juror receives a utility of 0 for a decision that matches the state of the world, $-\beta^i$ if the jury convicts the innocent, and $-(1-\beta^i)$ if the jury acquits the guilty. As standard, we assume that $\beta^i \in (0.5, 1)$.

We assume that to convict, a unanimous vote for conviction must be reached. As all information is shared, there are no "pivotal" considerations, so we can simply assume that each juror votes sincerely.[26] That is, she votes to acquit (convict) if she prefers to do so given her post-communication beliefs and her preferences.

[24] Note that we can extend the above analysis to include prices determined by 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). See also the surveys of Vayanos and Wang (2013) and Bikchandani and Sharma (2000).

[25] Goeree and Yariv (2011) and Guarnaschelli et al (2000) have shown that when the option to deliberate (or to conduct a straw poll before decisions) is offered, individuals overwhelmingly tend to be truthful, even when there are conflicting preferences.

[26] The analysis can then be easily extended to other voting rules.

Let $\hat{q}$ be the number such that $\frac{\hat{q}}{1-\hat{q}}$ is the geometric average of $\{\frac{q^j}{(1-q^j)}\}_{j\in N}$. As a juror uses her *worst-case* utility to evaluate her payoff from convicting versus acquitting, she convicts if and only if:

$$\begin{aligned}
\max_{\eta^i(I)\in C(a^i,\mathbf{q}_n)} \eta^i(I)(-\beta^i) &> \max_{\eta^i(G)\in C(a^i,\mathbf{q}_n)} \eta^i(G)(-(1-\beta^i)) \Leftrightarrow \\
\frac{\beta^i}{1-\beta^i} &< \frac{\max_{\eta^i(G)\in C(a^i,\mathbf{q}_n)} \eta^i(G)}{1-\min_{\eta^i(G)\in C(a^i,\mathbf{q}_n)} \eta^i(G)} \Leftrightarrow \\
\frac{\beta^i}{1-\beta^i} &< \frac{a^i+\frac{1}{a^i}(\frac{\hat{q}}{1-\hat{q}})^n}{a^i+\frac{1}{a^i}(\frac{\hat{q}}{1-\hat{q}})^n}
\end{aligned}$$

Let $K(\beta^i,a^i) = \frac{1}{2}(\sqrt{4\frac{\beta^i}{1-\beta^i}+(a^i)^4(\frac{2\beta^i-1}{1-\beta^i})^2}+(a^i)^2\frac{2\beta^i-1}{1-\beta^i})$. It is easy to see that this is a function that increases in β^i and a^i, and satisfies $K(\beta^i,1)=\frac{\beta^i}{1-\beta^i}$. We then have:

Proposition 5: *(i) The decisive juror is juror $i^*\in N$ which maximizes $K(\beta^i,a^i)$ for all $i\in N$.*[27] *Hence the jury convicts (acquits) iff*

$$(\frac{\hat{q}}{1-\hat{q}})^n > K(\beta^{i*},a^{i*}).$$

(ii) As n grows large, if $\frac{\hat{q}}{1-\hat{q}}\geqslant\delta>1$ ($\frac{\hat{q}}{1-\hat{q}}\leqslant\delta<1$) the jury only convicts (acquits).

The result gives a simple characterisation of jury behaviour: the information held by the group indicating that the defendant is guilty (summarized by the geometric mean $\hat{q}_n$) has to be strong compared to the cutoff $K(\beta^{i*},a^{i*})$ of the decisive juror. The decisive juror is determined not only by her taste parameter β^{i*}, but also by her correlation parameter a^{i^*}. The less she considers correlation, the more "persuasive" is the information held by the group, while the more she considers correlation, the more cautious she will be.

Are large groups better or worse than small groups? Note that now we identify two obstacles for efficiency. First, as in the investment application, the cautiousness and the NB effect may impede the behaviour of each juror, the former when the group is small and the latter when the group is large. Second, in the juries case, the decisive juror is not necessarily the "right" one. Unanimity rule indicates that the decisive juror should be the one with the highest β^i, for any known joint information structure. However, in our model, the identity of the decisive juror will also depend on a^i; specifically, it could be that the most lenient juror will not consider sufficient correlation and hence will not be the decisive one.

[27] Specifically, $K(\beta^i,a^i)=\frac{1}{2}(\sqrt{4\frac{\beta^i}{1-\beta^i}+(a^i)^4(\frac{2\beta^i-1}{1-\beta^i})^2}+(a^i)^2\frac{2\beta^i-1}{1-\beta^i})$.

To consider efficiency we contrast two possible joint information structures: that of (conditional) independence and that of full correlation. As before, the interesting case to consider arises when $\frac{\hat{q}}{1-\hat{q}} > 1$.[28] In the result below we compare the equilibrium behaviour of juries to the efficient course of action, for each of these information structures.

Proposition 6: *(i) Suppose that the jurors receive conditionally independent signals on the state of the world. Then, when a^{i^*} is large enough, small juries will acquit too frequently (compared to the efficient course of action) while large enough juries will behave efficiently. (ii) Suppose that there is one draw of an informative signal which all jurors observe. Then small juries acquit too frequently for a large a^{i^*} and convict too frequently for a small a^{i^*}. When the jury is large it convicts too frequently.*

To see the intuition, note first that the optimal course of behaviour, given the social desire for unanimity, is to compare the cutoff for the most extreme judge with the aggregated information. That is, when information sources are independent, the jury should convict iff $\max_{i \in N} \beta^i < \frac{1+\frac{\hat{q}}{1-\hat{q}}^n}{2+(\frac{1-\hat{q}}{\hat{q}}^n+\frac{\hat{q}}{1-\hat{q}}^n)}$, and when information sources are fully correlated, then the jury should convict iff $\max_{i \in N} \beta^i < q$.

How does behaviour compare to the efficient one? One feature of inefficiency was already identified in the previous Section. Specifically, consider one juror's decision. We have seen that when $\frac{\hat{q}}{1-\hat{q}} > 1$, a small number of forecasts can induce a cautious behaviour (that is, in favour of acquittal), while a large number of forecasts will induce a behaviour in line with correlation neglect. When the signals are indeed independent, this means that small juries over-acquit, and that large juries behave efficiently.[29] When the signals are fully correlated, it is large juries that induce a conviction bias.[30]

[28] When $\frac{\hat{q}}{1-\hat{q}} < 1$, all individuals would acquit for all a^i (recall that $\beta^i > 0.5$) which is the efficient course of action for both information structures we consider.

[29] This is in contrast with the predictions of Feddersen and Pesendorfer (1998), but in line with the experiments of Goeree and Yariv (2011) and Guarnaschelli et al (2000). These experiments allow for deliberation and find that larger juries are more accurate in their decisions.

[30] The experiments reported in Schkade et al (2000) show a severity shift that arises when jurors deliberate so that they tend to award higher punitive damages. In Schkade et al (2000) there is little control over both the information sources of participants and what they might believe about them: Participants are only shown the same videos of evidence and are not given any additional information. Our model generates such shifts for large groups; for the case described above, some individuals with $\beta^i > q > \frac{1}{2}$ would acquit pre-deliberation, while all would convict post-deliberation.

However, in contrast to the previous Section, collective decision making introduces another inefficiency which arises as the decisive jury is not the "right" one; when the signals are correlated, small juries can also *over-convict* if a^{i^*} is too low. Intuitively, the decisive juror is determined both by her preference cutoff and correlation parameter, while efficiency demands that only the preference parameter, together with the true information structure, should determine who is decisive. Specifically, efficiency means that the jury should convict if $\max_{i \in N} \beta^i < q$ while in our model the jury convicts iff $\beta^{i^*} < \frac{a^{i^*} + \frac{1}{a^{i^*}} \frac{\hat{q}}{1-\hat{q}}^{n}}{2a^{i^*} + \frac{1}{a^{i^*}} (\frac{1-\hat{q}}{\hat{q}}^{n} + \frac{\hat{q}}{1-\hat{q}}^{n})}$. However, as $\max \beta^i \geq \beta^{i^*}$, and if the decisive juror does not perceive sufficient correlation, it can be that $\frac{a^{i^*} + \frac{1}{a^{i^*}} \frac{\hat{q}}{1-\hat{q}}^{n}}{2a^{i^*} + \frac{1}{a^{i^*}} (\frac{1-\hat{q}}{\hat{q}}^{n} + \frac{\hat{q}}{1-\hat{q}}^{n})} > \beta^{i^*}$ while $\max_{i \in N} \beta^i > q$.

5 Conclusion

We suggest a new framework to analyse how sophisticated decision makers make decisions when they face ambiguity over the correlation of multiple sources of information. The decision makers generate a set of predictions based on a set of "correlation scenarios" and take a decision based on their attitudes towards ambiguity. Their set of predictions are fully characterised by the level of correlation they consider and the Naïve-Bayes interpretation of the information. A larger consideration set of correlation scenarios increases ambiguity and therefore induces more conservative or cautious behaviour. On the other hand the level of information implicit in a Naïve-Bayes interpretation of forecasts pushes individuals or organisations to be more confident and sometimes engage in risky behaviour. We therefore uncover a relation between complexity, confidence, and correlation, and illustrate its effects on CDO rating, investment behaviour, and group decision making.

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] Bikhchandani, S. and S. Sharma (2000). Herd Behaviour in Financial Markets- A Review. IMF working paper WP/00/48.

[5] Bohren, A. (2016): "Informational Herding with Model Misspecification," Journal of Economic Theory, 222–247.

[6] Bose, S. and Renou, L. (2014). Mechanism Design with Ambiguous Communication Devices. Econometrica 82: 1853-1872.

[7] Brunnermeier M.K. (2009). Deciphering the Liquidity and Credit Crunch 2007-2008. Journal of Economic Perspectives, 23, 77–100.

[8] Church, K.W and P. Hanks (1991), Word association norms, mutual information, and lexicography. Comput. Linguist. 16 (1): 22–29.

[9] Coval, J., J. Jurek, and E. Stafford (2009). The Economics of Structured Finance. Journal of Economic Perspectives 23, 3–25.

[10] Dell, M. and P. Querbin (2017). Nation Building Through Foreign Intervention: Evidence from Discontinuities in Military Strategies. Working paper, Harvard and NBER.

[11] De Marzo PM, Vayanos D, and Zwiebel J. (2003). Persuasion bias, social influence and unidimensional opinions. Q. J. Econ. 118:909–68.

[12] Domingos, P., and M. J. Pazzani. 1996. Beyond Independence: conditions for the optimality of the simple Bayesian classifier. *Machine Learning: Proceedings of the Thirteenth International Conference*, L. Saitta, eds. : Morgan Kaufman.

[13] Duffie, D., A. Eckner, G. Horel, and L. Saita. (2009). Frailty Correlated Default. The Journal of Finance, Vol. LXIV, No. 5, October.

[14] Easley, D. and M. O'hara (2009), Ambiguity and Nonparticipation: The Role of Regulation, Review of Financial Studies 22, 1817-1843.

[15] Ellis, A. and M. Piccione. Correlation Misperception in Choice. Forthcoming. American Economic Review.

[16] Enke, B. and F. Zimmerman (2013). Correlation Neglect in Belief Formation. mimeo.

[17] Epstein, L. and M. Schneider. (2007). Learning under Ambiguity. Review of Economic Studies. 74, 1275–1303.

[18] Epstein, L. and Y. Halevy (2019), Ambiguous Correlation, The Review of Economic Studies, Volume 86, Issue 2, Pages 668–693.

[19] Esponda, I. and D. Pouzo (2016): "Berk-Nash Equilibrium: A Framework for Modeling Agents with Misspecified Models," Econometrica, 84, 1093–1130.

[20] Eyster, E. and G. Weizsäcker (2011). Correlation Neglect in Financial Decision-Making. Discussion Papers of DIW Berlin 1104.

[21] Gilboa, I. and D. Schmeidler (1989), Maxmin expected utility with non-unique prior, Journal of Mathematical Economics 18, 141-153.

[22] Glaeser, E. and Cass R. Sunstein (2009), Extremism and social learning." Journal of Legal Analysis, Volume 1(1), 262 - 324.

[23] Hanani, E. and P. Klibanoff (2007). Updating preferences with multiple priors. Theoretical Economics 2, 261-298.

[24] Heidhues, P., B. Koszegi, and P. Strack (2018): "Unrealistic Expectations and Misguided Learning," Econometrica.

[25] Hodges J. and L. Le Cam (1960). The Poisson Approximation to the Poisson Binomial Distribution. The Annals of Mathematical Statistics. 737-740.

[26] Jiang, J. and W. Tian (2016). Correlation Uncertainty, Heterogeneous Beliefs and Asset Prices. mimeo, University of North Carolina.

[27] Kallir, I. and Sonsino, D. (2009). The Perception of Correlation in Investment Decisions. Southern Economic Journal 75 (4): 1045-66.

[28] Laohakunakorn, K., G. Levy and R. Razin (forthcoming). Common Value Auctions with Ambiguity over Information Structures and Correlation Capacities. Journal of Economic Theory.

[29] Levy, G. and R. Razin. (2015a). Correlation Neglect, Voting Behaviour and Information Aggregation. American Economic Review 105: 1634-1645.

[30] 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.

[31] Li, David X. (2000). On Default Correlation: A Copula Function Approach. Journal of Fixed Income. 9 (4): 43–54.

[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] MacKenzie D, Spears T (2014). 'The formula that killed Wall Street': The Gaussian copula and modelling practices in investment banking. Social Studies of Science 44(3): 393–417

[34] Nelsen, R. (2006). An Introduction to Copulas. Springer.

[35] Ortoleva, P. and E. Snowberg. (2015). Overconfidence in political economy. *American Economic Review,* 105: 504-535.

[36] Querubin, P. and M. Dell, M. (2016): "Nation Building Through Foreign Intervention: Evidence from Discontinuities in Military Strategies," mimeo.

[37] Russell, Stuart; Norvig, Peter (2003). *Artificial Intelligence: A Modern Approach* (2nd ed.). Prentice Hall.

[38] 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.

[39] Schucany, W.R., Parr, W.C., Boyer, J.E.: Correlation structure in Farlie–Gumbel–Morgenstern distributions. Biometrika 65, 650–653 (1978)

[40] Seidenfeld, T. and Wasserman, L. (1993), Dilation for Sets of Probabilities. Annals of Statistics. 21 no. 3, 1139–1154.

[41] Siniscalchi, M. (2011). Dynamic choice under ambiguity. Theoretical Economics 6, 379–421.

[42] Sobel, J. (2014). On the relationship between individual and group decisions, Theoretical Economics 9, 163–185.

[43] 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.

[44] 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.

6 Appendix

6.1 Appendix A

Proof of Proposition 1.

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

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

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

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

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

Construct the new information structure $(S', f'(.|\omega))$ by keeping the same distribution over signals as in (S, f), while keeping the label s^* and bundling all possible signals $s \neq s^*$

under one signal s^{-*}. In particular, $\forall \omega \in \Omega$,

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

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

It remains to show that the ePMI constraints hold for (S', f') so that it satisfies A1. Note first that the ePMI constraint for $(s^*, ..., s^*)$ holds by definition of (S', f'). 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, f) as $\frac{\sum_{l=1}^m c_l}{\sum_{l=1}^m c'_l}$ where $c_l = f(\mathbf{s}_l|\omega)$ for some $\mathbf{s}_l = (s_l^1, ..., s_l^k) \in S$ where we sum over all $\mathbf{s}^l$ that compose $\mathbf{s}$, and $c'_l = \prod_{j \in K} f^j(s_l^j|\omega)$. 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', f').■

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 Proposition, 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 ω 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 ω, 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 ω. Consider first the event in which only one source had received s^*.

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

The ePMI is:

$$\frac{\varepsilon\frac{q^j(\omega)}{p(\omega)} - \alpha_\omega - \varepsilon\frac{q^j(\omega)}{p(\omega)}(\sum_{\substack{M \subset K/\{j\} \\ |M| \geq 1}} \prod_{i \in M} \delta_\omega^i \prod_{l \in K/M \cup \{j\}} (1 - \delta_\omega^l))}{\varepsilon\frac{q^j(\omega)}{p(\omega)} \prod_{l \neq j}(1 - \varepsilon\frac{q^l(\omega)}{p(\omega)})}$$

$$= \frac{1 - \lambda_\omega \prod_{l \in K/j} (\varepsilon\frac{q^l(\omega)}{p(\omega)}) - (\sum_{\substack{M \subset K/\{j\} \\ |M| \geq 1}} \prod_{i \in M} \delta_\omega^i \prod_{l \in K/M \cup \{j\}} (1 - \delta_\omega^l))}{\prod_{l \neq j}(1 - \varepsilon\frac{q^l(\omega)}{p(\omega)})} \rightarrow_{\varepsilon \to 0} 1,$$

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

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

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

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:

$$\frac{(1-\varepsilon\frac{q^j(\omega)}{p(\omega)})}{(1-\varepsilon\frac{q^j(\omega)}{p(\omega)})\prod_{l\neq j}(1-\varepsilon\frac{q^j(\omega)}{p(\omega)})}$$

$$-\frac{(1-\varepsilon\frac{q^j(\omega)}{p(\omega)})(\sum_{\substack{M\subset K/\{j\}\\|M|\geq 2}}\prod_{i\in M}\delta^i_\omega\prod_{l\in K/M\cup\{j\}}(1-\delta^l_\omega))-(k-1)(\varepsilon\frac{q^j(\omega)}{p(\omega)}-\alpha_\omega-\varepsilon\frac{q^j(\omega)}{p(\omega)}(\sum_{\substack{M\subset K/\{j\}\\|M|\geq 1}}\prod_{i\in M}\delta^i_\omega\prod_{l\in K/M\cup\{j\}}(1-\delta^l_\omega)))}{(1-\varepsilon\frac{q^j(\omega)}{p(\omega)})\prod_{l\neq j}(1-\varepsilon\frac{q^j(\omega)}{p(\omega)})}$$

$$= \frac{1-(\sum_{\substack{M\subset K/\{j\}\\|M|\geq 2}}\prod_{i\in M}\delta^i_\omega\prod_{l\in K/M\cup\{j\}}(1-\delta^l_\omega))}{\prod_{l\neq j}(1-\varepsilon\frac{q^j(\omega)}{p(\omega)})}$$

$$-\varepsilon\frac{(k-1)(\frac{q^j(\omega)}{p(\omega)}-\lambda_\omega\varepsilon^{k-1}\prod_{i\in K}(\frac{q^i(\omega)}{p(\omega)})-\frac{q^j(\omega)}{p(\omega)}(\sum_{\substack{M\subset K/\{j\}\\|M|\geq 1}}\prod_{i\in M}\delta^i_\omega\prod_{l\in K/M\cup\{j\}}(1-\delta^l_\omega)))}{(1-\varepsilon\frac{q^j(\omega)}{p(\omega)})\prod_{l\neq j}(1-\varepsilon\frac{q^j(\omega)}{p(\omega)})}$$

$$\rightarrow_{\varepsilon\rightarrow 0} 1.$$

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

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

Consider the belief $\eta(\boldsymbol{\omega}) = \frac{\lambda_\omega \eta^{NB}(\boldsymbol{\omega})}{\sum_v \lambda_\mathbf{v}\eta^{NB}(\mathbf{v})}$. Let $f(s,\boldsymbol{\omega}) = \prod_{j\in K} f^j_i(s^j_i|\omega_i)$. Let $f^j_i(s^j_i|\omega_i) = \frac{q^j_i(\omega_i)}{p_i(\omega_i)}$ and let $p(\boldsymbol{\omega}) = \lambda_{\boldsymbol{\omega}}\prod_{i\in N} p_i(\omega_i)$. The ePMI constraints are satisfied as well as the rationalisability constraints as $\frac{p_i(\omega_i)f^j_i(s^j_i|\omega_i)}{\sum_{v_i}p_i(v_i)f^j_i(s^j_i|v_i)} = \frac{q^j_i(\omega_i)}{\sum_{v_i}q^j_i(v_i)} = q^j_i(\omega_i)$. Moreover the belief can be generated by $\eta(\boldsymbol{\omega}) = \frac{p(\boldsymbol{\omega})\prod_{i\in N}\prod_{j\in K_i} f^j_i(s^j_i|\omega_i)}{\sum_\mathbf{v} p(\mathbf{v})\prod_{i\in N}\prod_{j\in K_i} f^j_i(s^j_i|v_i)} = \frac{p(\boldsymbol{\omega})\prod_{i\in N}\prod_{j\in K_i} f^j_i(s^j_i|\omega_i)}{\sum_\mathbf{v} p(\mathbf{v})\prod_{i\in N}\prod_{j\in K_i} f^j_i(s^j_i|v_i)} = \frac{\lambda_{\boldsymbol{\omega}}\eta^{NB}(\boldsymbol{\omega})}{\sum_v \lambda_\mathbf{v}\eta^{NB}(\mathbf{v})}$ 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 η and η' that are in $C(a,\mathbf{q})$. Note that from the above a belief $\eta(.)$ is in $C(a,\mathbf{q})$ if and only if for any $v,\omega\in\Omega$ we have,

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

Thus all likelihood ratios satisfy,

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

To prove convexity we show that we can find a vector $\boldsymbol{\lambda}^{\beta}$ 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 η,η' satisfy

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

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,

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

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

Proof for Example 2: It is easy to see that whenever a share of the realisations of the same posterior is not qK, then the PMI equals $1-\alpha$. The largest ePMI arises then when the state is $\omega=1$ (0) and share of some specific qK realisations is q $(1-q)$ and the other q $(1-q)$. This is therefore the upper bound of a_K, and it is

$$\frac{\alpha\frac{1}{\frac{K!}{(qK)!((1-q)K)!}}+(1-\alpha)q^{qK}(1-q)^{(1-q)K}}{q^{qK}(1-q)^{(1-q)K}}$$

where in the nominator we have the probability of this event arising under the information structure considered, and in the denominator the probability of this event arising under independence, using the marginal probabilities. Note that a_K will increase in the slowest rate as it is largest under the event which is also the most probable under independence (for large K).

Re-arranging and using Stirling's formula for $\frac{K!}{(qK)!((1-q)K)!}$, this becomes

$$\begin{aligned}&\alpha\frac{1}{\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{Kq(1-q)}(q)^{qK}(1-q)^{(1-q)K}}q^{qK}(1-q)^{(1-q)K}}+(1-\alpha)\\ =\;&\alpha\sqrt{2\pi}\sqrt{Kq(1-q)}+(1-\alpha)\rightarrow_{K\rightarrow\infty}\infty\end{aligned}$$

Let us consider the event in which all forecasts are q. In this case, $\eta^{NB}_{\mathbf{q}_K}(1)=\frac{q^K}{q^K+(1-q)^K}$. Our limit condition demands

$$\lim_{K\rightarrow\infty}a_K^2(1-\eta^{NB}_{\mathbf{q}_K}(1))=0$$

Note that

$$\begin{aligned}
& a_K^2(1-\eta_K^{NB}(1)) \\
= \; & (\alpha\sqrt{2\pi}\sqrt{Kq(1-q)}+(1-\alpha))^2\frac{(1-q)^K}{q^K+(1-q)^K} \\
= \; & (\alpha^2 2\pi Kq(1-q)+(1-\alpha)^2+(1-\alpha)\alpha\sqrt{2\pi}\sqrt{Kq(1-q)})\frac{(1-q)^K}{q^K+(1-q)^K}
\end{aligned}$$

It is therefore sufficient to show that $\lim_{K\to\infty}\alpha^2 2\pi Kq(1-q)\frac{(1-q)^K}{q^K+(1-q)^K}=0$. But note that

$$\lim_{K\to\infty} K\frac{(1-q)^K}{q^K+(1-q)^K}=\lim_{K\to\infty} K\frac{\frac{(1-q)^K}{q^K}}{1+\frac{(1-q)^K}{q^K}}<\lim_{K\to\infty} K\frac{(1-q)^K}{q^K}=0.\blacksquare$$

Proof of Remark 1: 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:

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

which equals the formulation in the text.$\blacksquare$

Proof of Proposition 2: By the approximation and from Remark 1, we know that the worst case scenario is

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

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

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

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

Proof of Proposition 3: In text.

Proof of Proposition 4: (i) The proof follows from the construction in Proposition 1, Observation 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 Proposition 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 Proposition 5: In text.

Proof of Proposition 6: Consider the case of independent information. Note that the optimal course of behaviour is to convict iff $\max_{i\in N} \beta^i < \frac{1+\frac{\hat{q}}{1-\hat{q}}^n}{2+(\frac{1-\hat{q}}{\hat{q}}^n+\frac{\hat{q}}{1-\hat{q}}^n)}$. As for all $a^{i^*} > 1$, $\frac{a^{i^*}+\frac{1}{a^{i^*}}\frac{\hat{q}}{1-\hat{q}}^n}{2a^{i^*}+\frac{1}{a^{i^*}}(\frac{1-\hat{q}}{\hat{q}}^n+\frac{\hat{q}}{1-\hat{q}}^n)} < \frac{1+\frac{\hat{q}}{1-\hat{q}}^n}{2+(\frac{1-\hat{q}}{\hat{q}}^n+\frac{\hat{q}}{1-\hat{q}}^n)}$, the only distortion arises when $\frac{a^{i^*}+\frac{1}{a^{i^*}}\frac{\hat{q}}{1-\hat{q}}^n}{2a^{i^*}+\frac{1}{a^{i^*}}(\frac{1-\hat{q}}{\hat{q}}^n+\frac{\hat{q}}{1-\hat{q}}^n)} \leq \beta^{i^*} \leq \max_{i\in N} \beta^i \leq \frac{1+\frac{\hat{q}}{1-\hat{q}}^n}{2+(\frac{1-\hat{q}}{\hat{q}}^n+\frac{\hat{q}}{1-\hat{q}}^n)}$, which implies that juries over-acquit. When n is large enough juries behave according to Naive-Bayes belief and hence no distortion arises.

Consider now the case in which there is one draw of an informative signal and all observe the same draw and form a posterior $q^i(G) = q$. Efficiency means that that the jury should convict if $\max_{i\in N} \beta^i < q$ while in our model the jury convicts iff $\beta^{i^*} < \frac{a^{i^*}+\frac{1}{a^{i^*}}\frac{\hat{q}}{1-\hat{q}}^n}{2a^{i^*}+\frac{1}{a^{i^*}}(\frac{1-\hat{q}}{\hat{q}}^n+\frac{\hat{q}}{1-\hat{q}}^n)}$. Fix n to be small enough. When a^{i^*} is large enough then $\frac{a^{i^*}+\frac{1}{a^{i^*}}\frac{\hat{q}}{1-\hat{q}}^n}{2a^{i^*}+\frac{1}{a^{i^*}}(\frac{1-\hat{q}}{\hat{q}}^n+\frac{\hat{q}}{1-\hat{q}}^n)} < q$. Otherwise $\frac{a^{i^*}+\frac{1}{a^{i^*}}\frac{\hat{q}}{1-\hat{q}}^n}{2a^{i^*}+\frac{1}{a^{i^*}}(\frac{1-\hat{q}}{\hat{q}}^n+\frac{\hat{q}}{1-\hat{q}}^n)} > q$. Thus, for a small n, if a^{i^*} is large we can have an acquittal bias as before (cautiousness effect), while for a small a^{i^*} we can have a conviction bias due to the neglect of correlation, in the case in which $\frac{a^{i^*}+\frac{1}{a^{i^*}}\frac{\hat{q}}{1-\hat{q}}^n}{2a^{i^*}+\frac{1}{a^{i^*}}(\frac{1-\hat{q}}{\hat{q}}^n+\frac{\hat{q}}{1-\hat{q}}^n)} > \max_{i\in N} \beta^i > q$. As

$\max_{i\in N}\beta^i \geq \beta^{i^*}$, the jury will convict while it is optimal to acquit. This effect arises because the decisive juror is not the right one. When n is high on the other hand, by the NB effect, $\frac{a^{i^*}+\frac{1}{a^{i^*}}\frac{\hat{q}}{1-\hat{q}}^n}{2a^{i^*}+\frac{1}{a^{i^*}}(\frac{1-\hat{q}}{\hat{q}}^n+\frac{\hat{q}}{1-\hat{q}}^n)} \to 1 > q$, and thus indeed we have a conviction bias when $\max_{i\in N}\beta^i > q$.■

6.2 Appendix B: Other results

6.2.1 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 ρ (Kendal's τ) in $[-\bar{\rho}, \bar{\rho}]$.*

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

Among all such information structures, take the supremum according to the highest copula. That information structure has a Spearman's ρ (Kendall's τ) 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 ρ (or Kendall's τ) 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 ρ (Kendall's τ) in between the two copulas above.■

6.2.2 Agents with private information

In the application in Section 4.2, 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 she ends up with a unique rationalised belief even though she can imagine many joint information structures. This we do in Proposition B2 below.

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 Proposition B3 we show that the results extend to the case in which the agent knows the marginal distributions and signals.

Proposition B2: *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 $f^j(\omega|s') = \frac{p(\omega)f^j(s'|\omega)}{\sum_{v\in\Omega} p(v)f^j(s'|v)}$.*

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 $(\times_{l=1}^k \hat{S}^l, \hat{f}(\mathbf{s}, \omega))$ for which $\sum_{\mathbf{s}^{-j}\in\times_{j\neq l}\hat{S}^l} \hat{f}(s', \mathbf{s}^{-j}|\omega) = f^j(s'|\omega)$ for all ω. For any such joint information structure $(\times_{l=1}^k \hat{S}^l, \hat{f}(\mathbf{s}, \omega))$, we generate the posterior belief about state ω as

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

for all ω.■

6.2.3 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. We now illustrate that relaxing these assumptions will not affect our qualitative results characterising $C(a, \mathbf{q})$; what does change is that the set of beliefs might depend on the particulars of these marginal information structures.

Consider for example the following information structure.[31] Assume that there are two information 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:

[31] As we show in the appendix, any information structure that rationalizes a set of forecasts can be replicated by a structure with two signals only for each forecaster. Thus, the example below, symmetry aside, is general for the case of two realisations of the state.

$\omega = 0$	s^*	s^{**}	$\omega = 1$	s^*	s^{**}
s^*	$f(s^*\|0) - f_0$	f_0	s^*	$f(s^*\|1) - f_1$	f_1
s^{**}	f_0	$1 - f(s^*\|0) - f_0$	s^{**}	f_1	$1 - f(s^*\|1) - f_1$

Note that even though the individual knows $f(s|\omega)$, she still does not know f_0 and f_1. Suppose now that she observes the forecasts as well as $f(s|\omega)$, which is equivalent to observing the signals and marginals. For expositional purposes, we focus on the case in which both information sources observed the signal s^*, i.e., $\mathbf{q} = (\frac{f(s^*|1)}{f(s^*|1)+f(s^*|0)}, \frac{f(s^*|1)}{f(s^*|1)+f(s^*|0)})$. We then have:

Proposition B3: *Given marginals $f(s^*|\omega)$ and forecasts $\mathbf{q} = (\frac{f(s^*|1)}{f(s^*|1)+f(s^*|0)}, \frac{f(s^*|1)}{f(s^*|1)+f(s^*|0)})$, the set of rationalisable beliefs is: (i) $C(a, \mathbf{q})$ as in Proposition 1 if $\frac{f(s^*|1)}{1-f(s^*|1)} \leq a \leq \frac{1-f(s^*|0)}{f(s^*|0)}$, (ii) Contained in $C(a, \mathbf{q})$, convex and contains the NB belief if $a < \frac{f(s^*|1)}{1-f(s^*|1)}$ or $\frac{1-f(s^*|0)}{f(s^*|0)} < a$.*

Proof: The ePMI constraints are:

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

First note that the Naïve-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 charecterising the highest and lowest values we can get for α_ω.

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

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

where the LHS inequality is derived from

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

Similarly,

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

where the RHS inequality is derived from

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

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

6.2.4 Continuous distributions: The FGM transformation

In this subsection we show that our analysis can be extended to continuous distributions. We use the FGM 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.[32]

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 marginals distributions are symmetric, $g(s^1|\omega)$ and $g(s^2|\omega)$, with PDFs $G(s^1|\omega)$ and $G(s^2|\omega)$ respectively. We assume that the agent perceives the following family of joint information structures, constructed according to the FGM 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:

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

It is then easy to show that Proposition 1 holds as well.[33] The set of rationalisable beliefs given some s, s' is the set of all beliefs $\eta(.|s, s')$ satisfying:

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

[32] In Laohakunakorn, Levy and Razin (forthcoming) we use this transformation to analyze the effects of corelation capacity on common value auctions.

[33] The sufficiency part is as in Section 3. The necessity part follows from directly from the assumption of the FGM family of functions.

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 Naïve-Bayes benchmark as before.