Wu DD, Olson DL. (2009). Enterprise risk management: Small business scorecards. *Production Planning and Control: The Management of Operations* 20(4): 362–369.
Young B. (2011). Accounting and risk management: The need for integration. *The Journal of Operational Risk* 6(1): 37–53.

7

Subjective Beliefs and Statistical Forecasts of Financial Risks: The Chief Risk Officer Project

Glenn W. Harrison and Richard D. Phillips

Introduction

Information about financial risks comes from many sources. We formally consider how one can elicit and use information from two important sources when making forecasts. One source is a traditional statistical forecast, using familiar econometric methods for extrapolating from the past to the future. The other source is the elicited subjective belief distributions of "experts" in this domain: Chief Risk Officers of major international corporations. We demonstrate how these beliefs can be elicited in a formal, structured and incentivized manner, and critically contain information on the precision of the individual's belief for each risk. We characterize the manner in which these two sources tell different stories about these risks, arguing that any distributional differences, or similarities, between the two sources are informative for risk managers. Finally, we characterize the degree of consistency among our experts: are they "on the same page" in their beliefs? We argue, again, that consistency or inconsistency of subjective beliefs is in itself informative for risk managers.[1]

We first explain the basic idea of bringing together these two sources of information on future risks, and what we seek to learn from this exercise. Then we review the subjective belief elicitation procedures we developed, and document the statistical model we developed to provide the traditional forecasts for comparison. The subsequent section offers some initial findings from our experts, and discusses how we characterize the consistency of beliefs: aggregate subjective beliefs and the statistical forecasts on the one hand, and the heterogeneous subjective beliefs themselves on the other hand. Finally, we conclude and discuss the possible extensions of our approach.

Complementary sources of information on financial risks

The motivation for our approach was very simple.[2] To us, it was apparent from the serious economic journalism of 2006, 2007 and early 2008 that there were many "canaries in the cave" that anticipated the landscape of the financial crisis of 2008. Of course, they may not have known precisely which domino might fall first, precisely when it might fall, how far it might fall, and which other dominoes it would take out with it. But just like the canary that is more sensitive to dangerous gas emissions than the human workers in the cave, and can alert them, albeit by dying, there were expert observers of the financial scene that anticipated some of the specific issues that actually occurred. To illustrate the accuracy of these forecasts, Figure 7.1 shows a selection of covers of *The Economist*. The top row all come from 2007. The second row all come from the early part of 2008, well before the crisis set in.

Who is it that is most likely to have "expert" views on the financial risks facing the global economy? Arguably, it is those that are paid well to provide advice on the management of those risks for large corporations, Chief Risk Officers (CROs). We therefore turn to this specific

Figure 7.1 The canaries in the cave

sample to provide us with subjective belief distributions about core financial risks.[3]

Following Savage (1971, 1972), we *define* subjective beliefs by the choices that individuals make when facing bets with outcomes that depend on those beliefs. To observe those choices, we conduct an experiment using proper scoring rules, which are simply structured bets offered to the individual by an observer (the experimenter). All of the elicited beliefs were incentivized and incentive-compatible, so that the CROs were making real choices with real economic consequences.

Our approach is to elicit the entire subjective belief *distribution* that an individual has, to ascertain how precise their knowledge is about a certain financial risk. The degree of precision of someone's belief is valuable information, hence professional risk managers spend time with professional forecasters at quarterly conferences, in order to better understand how firmly they hold to their predictions, and to hear the "back stories" that bring them closer to the conditionals underlying their forecasts. The importance of characterizing the distribution of beliefs has long been stressed in the forecasting literature, particularly by Zarnowitz and Lambros (1987), Manski (2004) and Engelberg et al. (2009).

A by-product of our characterization is that we can also say something about the degree of consistency in the subjective beliefs that a sample of CROs hold about a particular financial risk. It may be that the pooled belief distribution does not change from month to month, but underlying that stationary, pooled distribution are some individuals with significantly tighter beliefs and other individuals with significantly more diffuse beliefs. Those differences constitute valuable information, signaling that there is less consistency in the sample of experts than in the previous month, despite the pooled distribution being the same.[4]

We compare the subjective belief distributions we elicit with statistical forecasts coming from standard econometric models. We do not advocate our statistical forecasts as necessarily being better than those provided by professional forecasting firms, but ours do follow "state of the art" methods. Their purpose is to provide a transparent basis for evaluating the information content of the subjective beliefs. Are the canaries in the cave saying something other than the "objective indicators," perhaps because they are more sensitive? And if the subjective beliefs are consistent with the statistical forecasts, then we can presumably have greater confidence in both, implicitly pooling these information sources in a Bayesian manner.

The 11 financial risks we measure were selected to span equity risk, interest rate risk, currency risk, credit risk and commodity risk.[5] We define each index explicitly, which is common practice in prediction markets that allow real trades with real financial consequences:

- **The S&P 500 Index**. Standard and Poor's 500 Index is a capitalization-weighted index of 500 stocks. The index is designed to measure performance of the broad domestic economy through changes in the aggregate market value of 500 stocks, representing all major industries. The index was developed with a base level of ten for the 1941–1943 base period. The return does not include dividends paid, and is the final price divided by the starting price minus 1, quoted as a percentage. The *Bloomberg* terminal ticker symbol is SPX.
- **The Eurostoxx 50 (European Blue Chip, excluding the UK) Index**. This is a free-float market capitalization-weighted index of 50 European blue-chip stocks from those countries participating in the EMU. Each component's weight is capped at 10% of the index's total free float market capitalization. The index was developed with a base value of 1,000 as of December 31, 1991. The return does not include dividends paid, and is the final price divided by the starting price minus 1, quoted as a percentage. The *Bloomberg* terminal ticker symbol is SX5E.
- **The MSCI AC Asia (excluding Japan) Index.** This is a free-float weighted equity index. It was developed with a base value of 100 as of December 31, 1987. The return does not include dividends paid, and is the final price divided by the starting price minus 1, quoted as a percentage. The *Bloomberg* terminal ticker symbol is MXASJ.
- **The 10-Year US Treasury Bond Yield.** This is the yield to maturity of on-the-run 10-year United States Treasury Bonds. The *Bloomberg* terminal ticker symbol is GT10.
- **The 10-Year German Bund Yield.** This is the yield to maturity of on-the-run 10-year German Bund, which are government bonds. The *Bloomberg* terminal ticker symbol is GTDEM10TR.
- **The 10-Year Japanese Government Bond Yield.** This is the yield to maturity of on-the-run 10-year Japanese Government Bonds. The *Bloomberg* terminal ticker symbol is GJGB10.
- **The Euro/USD Exchange Rate.** Quoted as $ per €. The *Bloomberg* terminal ticker symbol is EURUSD.
- **The CDX North American Credit Default Swap Index.** The Markit CDX North America Investment Grade Index is composed of 125 equally weighted credit default swaps on investment grade entities, distributed among six sub-indices: High Volatility; Consumer; Energy; Financial; Industrial; and Technology, Media & Telecommunications. Markit CDX indices roll every six months in March & September. This is the quoted spread on the five-year basket credit derivative, with a coupon value of 100bps. The *Bloomberg* terminal ticker symbol is IBOXUMAE.
- **The iTraxx European Credit Default Swap Index.** The Markit iTraxx Europe Crossover index comprises 50 equally weighted credit default swaps on the most liquid sub-investment grade European corporate entities. The composition of each Markit iTraxx index is determined by a liquidity poll and certain criteria as determined by the index rules. The Markit iTraxx indices roll every six months in March and September. This is the quoted spread on the five-year basket credit derivative, with a coupon value of 500bps. The *Bloomberg* terminal ticker symbol is ITRXEXE.
- **Brent Crude Oil Price.** The price of current pipeline export quality Brent blend as supplied at Sullom Voe. The Inter Continental Exchange (ICE) Brent Futures is a deliverable contract based on Exchange of Futures for Physical (EFP) delivery with an option to cash settle. The contract price is in US dollars and cents per barrel. The *Bloomberg* terminal ticker symbol is CO1.
- **The Gold Spot Price** is quoted as US dollars per Troy Ounce. The *Bloomberg* terminal ticker symbol is GOLDS.

These indices span a range of the core financial risks affecting a wide range of global corporations.

There are many hypothetical surveys that elicit probabilistic forecasts for various events, where the term "probabilistic" is used in the general sense to refer to any attempt to elicit a *probability*, even if the entire distribution is not elicited. For instance, the most widely used subjective beliefs about longevity come from the *U.S. Health and Retirement Survey*, which has posed a simple question each year, since 1992, for respondents under the age of 65: "With 0 representing absolutely no chance, and 100 absolute certainty, what is the chance that you will live to be 75 years of age or older?" A comparable question asks the chance that they would live to be 85, and for respondents over 65, a variant asks the chances of them living 11–15 years more. Similar questions have been asked about returns to the S&P 500 in hypothetical surveys of Chief Financial Officers and US households by Graham and Harvey (2005) and Vissing-Jorgensen (2004), respectively.

There have also been many hypothetical surveys eliciting complete *distributions* over some event, reviewed by Manski (2004).[6] Prominent examples include the *U.S. Survey of Professional Forecasters* and beliefs about GDP and inflation, evaluated in Engelberg et al. (2009), and the *RAND American Life Panel Survey* and beliefs about inflation, evaluated in Bruin de Bruin et al. (2011).

Belief elicitation

We employ an explicit scoring rule to elicit reports that reveal the subjective belief distribution of the CRO. Figure 7.2 shows the general interface to elicit beliefs for one index, the S&P 500 Index. The key features of our approach are that individuals face incentives to truthfully reveal their entire subjective distribution. We do not just elicit subjective probabilities of events occurring, but whole distributions that reflect the confidence with which those events are expected. And we do not rely on hypothetical survey responses to encourage truthful and reflective responses.

The scoring rule

The decision maker in our experiment reports her subjective beliefs in a discrete version of a Quadratic Scoring Rule (QSR) for continuous distributions, developed by Matheson and Winkler (1976). Partition the domain into K intervals, and denote as r_k the report of the density in interval $k = 1, \ldots, K$. Assume for now that the decision maker is risk neutral, and that the full report consists of a series of reports for each interval, $\{ r_1, r_2, \ldots, r_k, \ldots, r_K \}$ such that $r_k \geq 0\ \forall k$ and $\Sigma_{i=1 \ldots K}(r_i) = 1$.

If k is the interval in which the true value lies, then the payoff score is from Matheson and Winkler (1976: 1088, equation 6):

$$S = (2 \times r_k) - \Sigma_{i=1 \ldots K}(r_i)^2$$

The reward in the score is a doubling of the report allocated to the true interval, and a penalty that depends on how these reports are distributed across the K intervals. The subject is rewarded for accuracy, but if that accuracy misses the true interval, the punishment is severe. The punishment includes all possible reports, including the correct one.

Consider some examples, assuming $K = 4$. What if the subject has very tight subjective beliefs and puts all of the tokens in the correct interval? Then the score is

$$S = (2 \times 1) - (1^2 + 0^2 + 0^2 + 0^2) = 2 - 1 = 1,$$

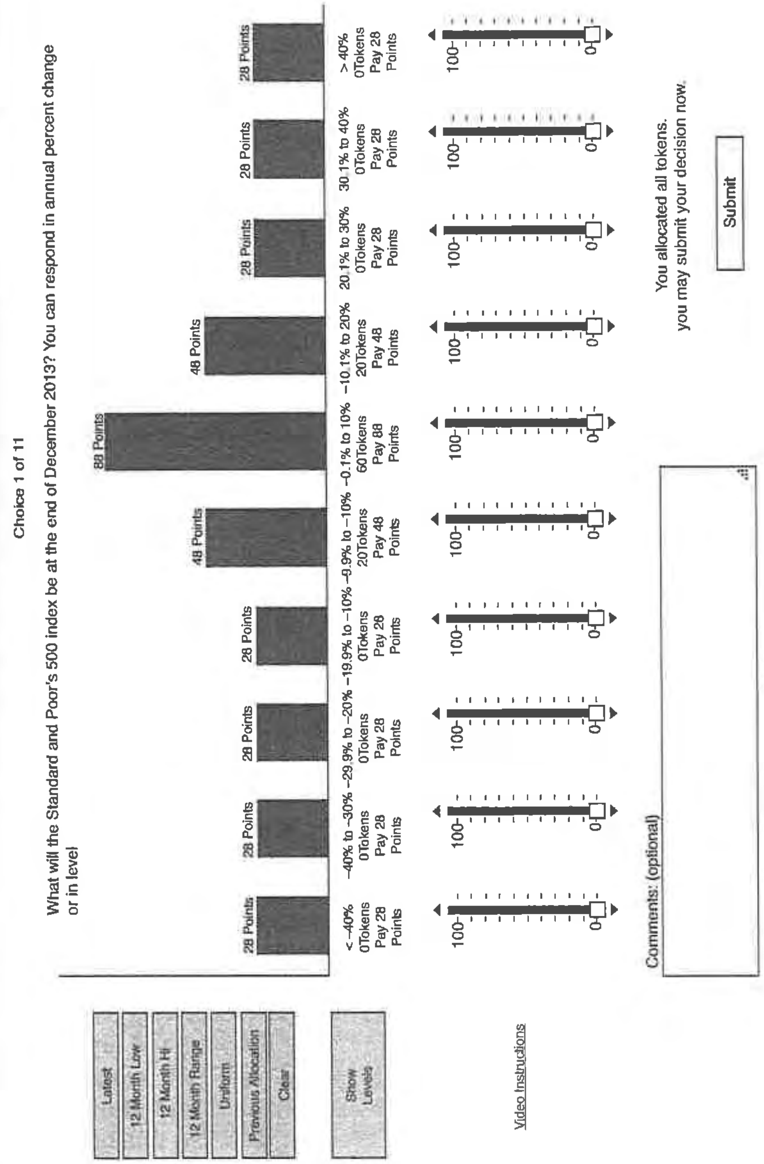


Figure 7.2 Belief elicitation interface

and this is positive. But if the subject has a tight subjective belief that is wrong, the score is

$$S = (2 \times 0) - (1^2 + 0^2 + 0^2 + 0^2) = 0 - 1 = -1,$$

and the score is negative. So we see that this score would have to include some additional "endowment" to ensure that the earnings are positive. Assuming that the subject has a very diffuse subjective belief and allocates 25% of the tokens to each interval, the score is less than 1:

$$S = (2 \times \tfrac{1}{4}) - (\tfrac{1}{4}^2 + \tfrac{1}{4}^2 + \tfrac{1}{4}^2 + \tfrac{1}{4}^2) = \tfrac{1}{2} - \tfrac{1}{4} = \tfrac{1}{4} < 1.$$

The trade-off from the last case is that one can always ensure a score of ¼, but there is an incentive to provide less diffuse reports, and that incentive is the possibility of a score of 1. To obtain the maximum score, one would have to re-allocate all tokens in Figure 7.2, for instance, to one interval, and then for that interval to the outcome.

To ensure complete generality, and avoid any decision maker facing losses, allow some endowment, α, and scaling of the score, β. We then get the generalized scoring rule:

$$\alpha + \beta\,[\,(2 \times r_k) - \Sigma_{i=1 \ldots K}\,(r_i)^2\,]$$

where we initially assumed α=0 and β=1. We can assume $\alpha > 0$ and $\beta \neq 0$ to get the payoffs to any level and units we want.

In our elicitation procedures $K = 10$, as shown in Figures 7.2 and 7.3, we do not know whether the subject is risk neutral. Indeed, the weight of evidence from past laboratory and field experiments clearly suggests that subjects will be modestly risk averse over the prizes they face (Harrison et al., 2007). It is well known that risk aversion can significantly affect inferences from applications of the QSR to eliciting subjective *probabilities* over *binary* events (Winkler and Murphy, 1970; Kadane and Winkler, 1988), and there are various methods for addressing these concerns.[7] Harrison et al. (2012) characterize the implications of the general case of a risk-averse agent when facing the QSR and reporting subjective *distributions* over *continuous* events, and find, remarkably, that these concerns do not apply with anything like the same force. For empirically plausible levels of risk aversion, one can reliably elicit the most important features of the latent subjective belief distribution without undertaking calibration for risk attitudes.

Specifically, Harrison et al. (2012) draw the following conclusions:

1. The individual never reports having a positive probability for an event that does not have positive subjective probability. So if the individual believes that the annual return on the S&P 500 Index is definitely below 20.1%, we would never see the individual reporting that it could be above 20.1%. Hence, we can infer from Figure 7.2, for instance, that this CRO truly attaches zero weight to this possibility, no matter what their risk attitudes.
2. If an individual has the same subjective probability for two events, then the reported probability will also be the same if the individual is risk averse or risk neutral. So if the individual attaches a true, subjective probability of 0.2 to the chance that the return on the S&P Index is between −9.9% and 0%, and a true, subjective probability of 0.2 to the chance that it is between 10.1% and 20%, the reported probabilities for these two intervals will be the same as well, as in Figure 7.2 (although, typically, not exactly 0.2, unless reporting a completely uniform distribution).
3. The converse is true for risk-averse subjects, as well as for risk lovers. That is, if we observe two events receiving the same reported probability, we know that the true probabilities are also equal, although not necessarily the same as the reported probabilities.
4. If the individual has a *symmetric* subjective distribution, then the reported mean will be *exactly* the same as the true subjective mean, whether or not the subjective distribution is unimodal. Hence, if we simply assume symmetry of the true distribution, a relatively weak assumption in many settings of interest, we can elicit the mean belief directly from the average of the reported distribution. So in the case

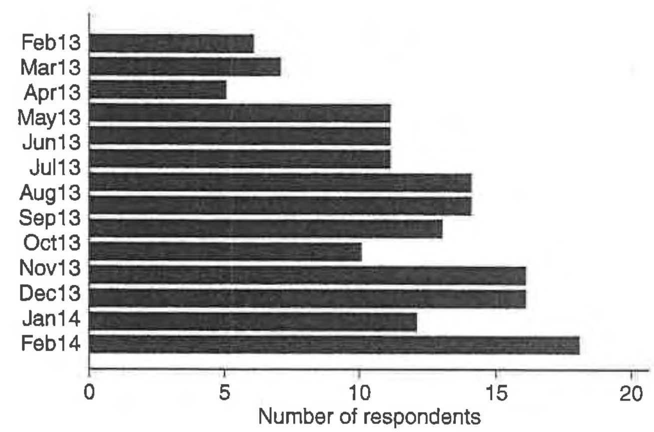


Figure 7.3 Number of CRO respondents by month

of the report in Figure 7.2, we know that the weighted average return of the reports, 5%, is in fact the average of the true subjective belief distribution.

5. The more risk averse an agent is, the more the reported distribution will resemble a uniform distribution defined on the support of their true distribution. In effect, risk aversion causes the individual to report a "flattened" version of their true distribution, but never to report beliefs to which they assign zero subjective probability. So if the reports in Figure 7.2 are from a risk-averse agent, we can infer that the true subjective beliefs place even *more* weight on the interval (0.1%, 10%) and *less* weight on the intervals (–9.9%, 0%) and (10.1%, 20%).
6. It is possible to capture the effect of increased risk aversion as the difference between the reported distribution and true distribution. This result provides a characterization of an empirical finding from incentivized experiments with objectively verifiable stimuli[8] that the reported distribution is "very close" to the true distribution for a wide range of empirically plausible risk attitudes. Harrison et al. (2012) show numerically that a priori plausible levels of risk aversion in laboratory and field settings imply no significant deviation between reported and true subjective beliefs in this setting.

Providing that our CROs exhibit the modest levels of risk aversion found universally, the lab and field settings for stakes of the level we used (e.g. Harrison and Rutström, 2008a), and make their choices solely in response to the incentives provided by the scoring rule, these results provide the basis for us using the reported distributions as if they are the true, subjective belief distributions.[9]

We went one step further and added a "binary lottery procedure," familiar from experimental economics, which was designed to encourage individuals to behave as if risk neutral. Hence, there is a reference in Figures 7.2 and 7.3 to payments in "points" rather than money. We explain this procedure below, and the evidence for its effects on behavior.

Incentives

The individuals we elicit beliefs from are all valuable employees of major corporations, and are compensated accordingly. Compensation packages for a CRO in top corporations are generally $1 million per year and above.[10] How does one incentivize such individuals to take the task seriously? We recognized that direct payments of the size we were able to make would not affect the pocket book of these individuals, so we decided to express the rewards instead as contributions to a charity that they selected from a list.[11] In effect, we are relying on these contributions to encourage our respondents to view their efforts as being compensated in the manner of a "gift exchange."[12]

We explained the incentive scheme in this manner in the video instructions:

> Research in behavioral economics has shown it is important that participants face an incentive scheme designed to reward them for taking the task seriously. The incentive mechanism we utilize is a way to convert the points you earn in the elicitation task into a chance of earning money towards a predefined charity of your choice. Every point you earn in a task gives you a greater chance of earning $50 for your selected charity. To be paid for this task, we will get a random number between 1 and 100, with each integer value equally likely. This random number will be the first and second decimals of the Dow Jones Industrial Index on the day we check your beliefs against the resulting value of the index. We use the DJIA to get a random number because it is easy to verify and public. So if the DJIA has a value of 12,649.35, the random number will be 35. We will round the DJIA to the nearest two decimal places, and a 00 (double zero) is interpreted as the random number 100. If this random number is less than or equal to your earned points, you earn $50, otherwise you earn $0.
>
> For instance, suppose you earned 79 points and then, using the DJIA decimals, the random number was 35. In this example, you would earn $50 since the random number is less than or equal to your earned points. However, if the random number was some number greater than 79, say 80, you earn $0. Additionally, note in cases where you earn 100 points, this will always result in a payment of $50, since every random number outcome from the DJIA decimals would result in a number less than or equal to 100. Thus, you can see that you will increase the chances of getting the $50 contribution to charity if you earned more points from your forecasting. This is the sole reason for this incentive scheme, to encourage you to earn more points.
>
> So it is up to you to balance the strength of your personal beliefs with the risk of them being wrong. There are two important points for you to keep in mind when placing your bets:
>
> 1. Your belief about the chances of each outcome is a personal judgment that depends on information you have about the different

events. Some people might be experts on a certain issue, and others might not be very knowledgeable about it. Your personal beliefs will naturally reflect your knowledge.
2. More points increase your chance of earning $50 for your charity. The points you earn will be compared with the random outcome from the DJIA decimals to determine whether your charity wins $50 or $0.

The key step in these procedures is the manner in which accurate reports are converted into the increased probability of a contribution to charity. We return to the logic and evidence of that step below (the instructional video shown later explained the manner in which payments would be made):

> The final item to cover in this training video is to describe how your earnings will be determined. As I mentioned earlier, all payoffs will be in the form of a donation to a registered charity that you should select from the list provided, after you go through the online trainer and complete a brief demographic survey. If you wish, we will make the donation in your name, and the charity will send you a letter acknowledging that for tax purposes. If you prefer to make the donation anonymously, that is fine as well. We will send you a copy of our payment of the donation made from your earnings.
>
> Each month you will be rewarded for accuracy on one of the 11 core risks randomly selected for you. Thus, the risk on which you are rewarded in one year's time is independent of other respondents, and it will change from month to month. As described earlier in this video, the amount of the payoff will be either $50 or $0 and will depend upon two factors:
>
> 1. the number of points you allocated to the interval that contains the true answer, which is determined one year later; and
> 2. based upon the first two decimal points of the closing value of the Dow Jones Industrial Average.
>
> The values of all indices will be determined as the closing value of the last trading day of the month one year later, as reported on *Bloomberg*.
>
> An account will be established for each participant to aggregate your monthly contributions to your preselected charity from July through June of each year, and the balance will be paid to that charity following the June contribution.

The report distribution chosen by the CRO only ever generates a probability of there being a $50 payment to charity instead of there being none at all.[13]

To see the logic of this procedure, and why it removes the effect of risk aversion, normalize the utility of the individual of the payment of $50 to 1, and the utility from the payment of $0 to 0. It is then immediate that the subject has had a linear utility function of money induced, as shown by Smith (1961). Given the theoretical results referred to earlier, we therefore predict that the individuals facing this elicitation mechanism will behave identically to those facing direct monetary payoffs. In fact, Harrison et al. (2012) show that this theoretical prediction is supported empirically in controlled experiments.[14]

All of these steps to ensure that there were some financial incentives, and that these were linked in a salient manner to the responses in the scoring rule, might seem elaborate.[15] Why not just ask a hypothetical survey question, and be done with it? The reason is that there are decades of evidence that subjects generally exhibit varying degrees of hypothetical bias in a wide range of elicitation tasks (e.g. Harrison and Rutström, 2008b). Delavande et al. (2011: 156) make the case for not bothering about incentives. Referring to studies in developing countries that have all been hypothetical, they argue that "even without payment, the answers received from such questions appear reasonable, and as such, there seems to have been a *de facto* decision that payments are not needed." We do not know what "reasonable" might possibly mean when it comes to subjective beliefs. In some settings, such as stated beliefs about longevity (e.g. Perozek, 2008), the metric for reasonableness appears to be whether the beliefs are actuarially correct on average. Although that is certainly of great policy interest, it is hard to know why it would be a metric for evaluating the validity of responses as reflecting the true subjective beliefs of individuals. Our preference is to build in some incentives for truthful responses, rather than theologically hope that incentives are not needed. The available evidence from convenience samples concludes that there are significant differences in elicited belief distributions between incentivized scoring rules and hypothetical surveys (Harrison, 2014). We welcome rigorous research into the reliability of hypothetical surveys of subjective beliefs for experts and the wider population.

We explicitly avoid having any competition or "tournament" between our respondents. This might appear superficially attractive as a way to motivate, but can quickly distort incentives for truthful reporting. For instance, imagine a setting in which one respondent needs a big score to improve his rank to be #1. Akin to a professional golfer that only cares about winning, and not coming second, one might expect extreme choices in an attempt to improve the ranking. Lichtendahl and Winkler (2007) formalize this intuition, and show that competitive incentives of this kind can turn a well-calibrated forecaster into someone that would appear overconfident of their beliefs. They also consider joint scoring rules that might mitigate this tendency, but we prefer to just avoid it by not using competitive incentives.

Measuring agreement and disagreement

Any measuring instrument can be compared against another one. Examples include weight scales, political opinion polls, or medical judgments about diagnoses. In our case, we are interested in the subjective beliefs about some fact, the value of some objective financial risk in a year, and seek to measure their consistency at a point in time.[16] In biostatistics literature, a popular concordance index ρ_c has been developed by Lin (1989, 2000). This index combines the familiar notion of correlation from a Pearson inter-class correlation coefficient with allowance for bias, and is virtually identical to measures of intra-class correlation used in psychology (Nickerson, 1997). The index is bounded between ± 1, with the usual interpretation that $\rho_c = 1$ indicates perfect concordance, and smaller values indicate poorer concordance.

We apply the concordance index in two ways. The first is to evaluate the consistency of the pooled subjective belief distribution over all respondents and the predictive distribution from the statistical model forecast. The second is to assess how much consistency there is across the different elicited subjective distributions of respondents.

There is a large literature on the significance of disagreement *across* elicited *point* forecasts as a measure of uncertainty *in* the forecast. These are different things, forced together solely because elicited *distributions* have not been available. Zarnowitz and Lambros (1987: 591ff.) pose the problem well:

> Although all forecasts are by their very nature probabilistic statements, most economic predictions quote but a single value to be assumed by a certain variable, without specifying the attached probabilities. Often many such point forecasts are available for a given target variable from a business outlook survey. If they show a high degree of agreement, does this indicate that the forecasters confidently expect the outcome they commonly predict to come true? More generally, does the dispersion of the point forecasts reflect their authors' uncertainty (i.e. their relative lack of confidence)?

Of course, it is perfectly possible for two uncertain subjective belief distributions to be in perfect agreement with each other. We therefore prefer the option of collecting richer, probabilistic data on forecasts, rather than searching for one or other modeling proxies for forecast uncertainty and disagreement (e.g. Lahiri and Sheng, 2010). In effect, this is an extension of the approach advocated by Zarnowitz and Lambros (1987: 618):

> We define "consensus" as the degree of agreement among corresponding point predictions by different individuals and "uncertainty" as the diffuseness of the probability distributions attached by the same individuals to their predictions.

We simply extend the concept of "consensus" to reflect the whole distribution, in the form of our measures of concordance.[17]

The elicitation interface

There are several additional features of our elicitation interface, shown in Figure 7.2. Each CRO selects an allocation of 100 tokens by sliding a bar for each bin, with the "histogram" representation changing in real time. Only when 100 tokens have been allocated can the allocation be submitted, and even then, there is a need to actively confirm the choice. This design extends the binary QSR interface single-slider developed by Andersen et al. (2014), which allows the experimenter to use a specific QSR to generate the implied allocations without burdening the CRO with messy formulae. Although our CRO subjects are all highly skilled applied mathematicians, dealing with formulae for things of this kind could easily distract from the main task at hand, representing one's beliefs.

A link to the video instructions explains the interface necessary for making choices. Although we formally verify that everyone has completed the training stage, because the instructions are delivered on the web there is no assurance that the training video was watched, or watched closely. Hence, it is useful to have this available as needed.

For several of the indices, such as the S&P 500 Index, it may be more natural for some respondents to consider their beliefs denominated in levels rather than annual percent returns. Hence, we provide an option for those subjects to instantly convert all labels to levels rather than percent returns, or back again if they choose. We are very careful to convert responses back to some common basis when reporting results, but it makes sense to allow respondents to choose the metric that is more natural to them.

Each interface also allows the respondent to initialize the distribution to reflect recent historical data or the last pooled subjective belief distribution. This is effected by "preset" tabs to the left of the interface. The allocation is always initialized at 0 tokens for every interval, but it is a simple matter to initialize a uniform distribution over all ten intervals, the historical range as a uniform distribution, the most optimistic or pessimistic historical outcome, or the latest historical value. We observe in pilots that the historical uniform initialization is a popular starting point.

Statistical forecasts

To provide some basis for judging the information content of the elicited subjective belief distributions, we generated statistical forecasts for the 11 risk indices for the same one-year horizon. We deliberately used transparent, familiar, "state of the art" statistical methods for these forecasts, since our objective was not to propose some novel statistical forecasting methodology.

The statistical models used were factor-augmented Vector AutoRegressions (VAR). The VAR model captures linear correlations between multiple economic time series, and is widely employed for forecasting financial indices such as these. The popularity of VAR models can be attributed to the fact that they often provide superior forecasts to those from univariate time-series models and structural simultaneous equations models. Indeed, Sims (1980) strongly advocates VAR models as providing a method to estimate economic relationships without the "incredible identification restrictions" of available structural models. Obviously, the forecasting ability of a model can suffer if false restrictions are imposed on it, and the domain in which we are forecasting is certainly one that does not have a settled structural model to be applied.

The VAR model is a natural generalization of the univariate autoregressive model to dynamic multivariate time series. A univariate autoregression is a single-equation, single-variable linear model in which the current value of a variable is determined by its own lagged values. In a VAR model, all variables are treated symmetrically so that each variable has an equation describing its evolutions over time based on its own lags and the lags of all the other variables appearing in the model. This simple framework provides a systematic way to capture rich dynamics in multiple time series, and the statistical VAR methodology is easy to use and interpret. Hamilton (1994) provides a detailed presentation of the statistical methodology for VAR analyses.

The factors of the *factor-augmented* VAR model are simply additional explanatory variables included along with the set of 11 variables to be forecast. The factor-augmented VAR is an approach advocated by Bernanke et al. (2005) for incorporating a broad range of conditioning information in otherwise standard VAR analyses. In choosing factors, we balanced parsimony and sophistication, and surveyed the available academic literature to identify those factors that had predictive value for our forecasting targets.

The parameters of the VAR models we use are estimated using time series of monthly observations. The estimated models are then used to produce 12-months ahead forecasts of the variables of interest by standard methods. A non-parametric bootstrap procedure is used to obtain joint predictive distributions. The bootstrap procedure used to construct predictive distributions follows the procedures developed by Thombs and Schucany (1990) for univariate autoregressive models and its extension by Kim (1999) to VAR models. This bootstrap procedure is particularly useful for forecasting purposes because it allows the construction of predictive distributions without assuming any particular distribution for the VAR model disturbances, and incorporates the effects of parameter uncertainty.

Initial results

The elicitation and forecasting activity described above has been implemented since January 2013. We document initial results, the manner in which we characterize results, and the nature of insights obtained.

Recruiting Chief Risk Officers

The experts in our subjective elicitation were recruited to join The Risk Council of *The Georgia State University CRO Risk Index* (http://www.gsucroriskindex.org/). The Risk Council consists of the risk professionals that participate in the monthly elicitation, and membership is limited to senior risk professionals.[18] By limiting participation to risk managers, we solicited the opinions of highly skilled professionals explicitly charged with forming opinions about the risks their firms face, but who themselves are not allowed to personally participate in markets.

The required duties of members of the Risk Council involve participating in the monthly elicitation. We designed the system recognizing the limited time that senior executives can allocate to this task. The web-based elicitation tool is designed so that users should be able to complete the monthly tasks within 15 minutes.

Risk Council members receive several benefits, apart from the incentives for charitable contributions built into the elicitation procedure itself. Participants are also entitled to a free subscription access to an anonymous version of the individual response data, and networking opportunities with other participants at optional round table events.

The recruitment of a CRO from a major corporation is a labor-intensive and network-intensive activity. We contacted potential respondents and explained the nature of the exercise, and generally received a very positive response from those we spoke to. Many then needed to obtain "legal" approval to participate, which is to be expected, despite the confidential nature of the responses. Every respondent had the option to identify themselves and their responses, but the default was to reveal only anonymous responses. The clear majority chose to keep their individual responses anonymous.

Figure 7.3 shows the steady increase in enrolment during the first 12 months. Our target after one year was roughly 30, and we anticipate being close to that number in several months. It is important to stress that we view this as a sample of experts in risk assessment, not as a sample of the general population.

Results

We present results based on the elicitation conducted in November 2013, with 16 CRO respondents. Figure 7.4 details the elicited individual distributions of three CRO respondents, identified by an anonymous number, for the S&P 500 Index, along with the pooled distribution over all CRO respondents. Figure 7.5 displays the elicited distributions for each CRO for the same equity risk. Table 7.1 summarizes the main findings for the elicited subjective beliefs and the statistical model used as a reference distribution. Figures 7.6 through 7.9 show the comparison of the statistical forecast and pooled subjective beliefs for each risk. Figure 7.6 displays results for the three equity indices, Figure 7.7 shows the three interest rates, Figure 7.8 shows the three financial indices and

Figure 7.9 shows the two commodity prices. For illustrative purposes, Figure 7.10 displays longitudinal results between February 2013 and November 2013 for each risk.

CRO beliefs indicate a low return environment for equities over the next 12 months. The mean forecasts are +1% for the US, 0% for Europe and +2% for Asia. Forecast levels in Europe and the US have moved upward with stock prices over the course of the year, but continue to indicate modest year-over-year changes. Although forecasts of mean returns are below historical averages, assessments of downside risks remain relatively low. The pooled belief distribution indicates a 6% chance of a decline of 20% or more in the S&P 500, which is roughly in line with historical averages. The assessment for Europe rests at 2%. The perceived risk in Asia is now 6%. The concordance indices in Table 7.1 suggest general agreement between the different CRO respondents about each of these equity risks. The concordance indices in Figure 7.6 suggest far less agreement between the two modeling approaches (subjective beliefs and the statistical model) with respect to European and Asian equities.

Since the first elicitation ten months ago, CRO forecasts have consistently indicated rising sovereign bond yields. The pooled subjective belief distribution in November 2013 indicates a 79% chance of an increase in the US Treasury yield over the next year. The corresponding assessments for the Bund and the JGB are 52% and 90%, respectively. Perceived risks of large yield movements are rising in relation to earlier forecasts. The pooled belief distribution indicates a 37% chance of a 50 basis point increase in the ten-year US Treasury yield. The corresponding figures for the Bund and JGB are 25% and 18%, respectively. The subjective figures for the Bund and the JGB are both at, or near, all-time highs since the first elicitation in February 2013. Table 7.1 also shows that the agreement across the *individual* subjective beliefs is much greater for US bond yields than for German or Japanese bond yields. Figure 7.7 shows that there is considerable disagreement between the two modeling approaches with respect to US and Japanese rates.

The column labeled "Value on 11/14/13" shows the value of each market risk at the close of business on the day before the elicitation. The statistics under the column labeled "Forecast Index" are the expected value and the standard deviation of the one-year-ahead forecast distribution for each risk estimated using the subjectively elicited and statistical distributions. The statistics under the column labeled "Percentage" show the expected percentage return and the standard deviation of returns for each risk. The statistics under the column labeled "Probability

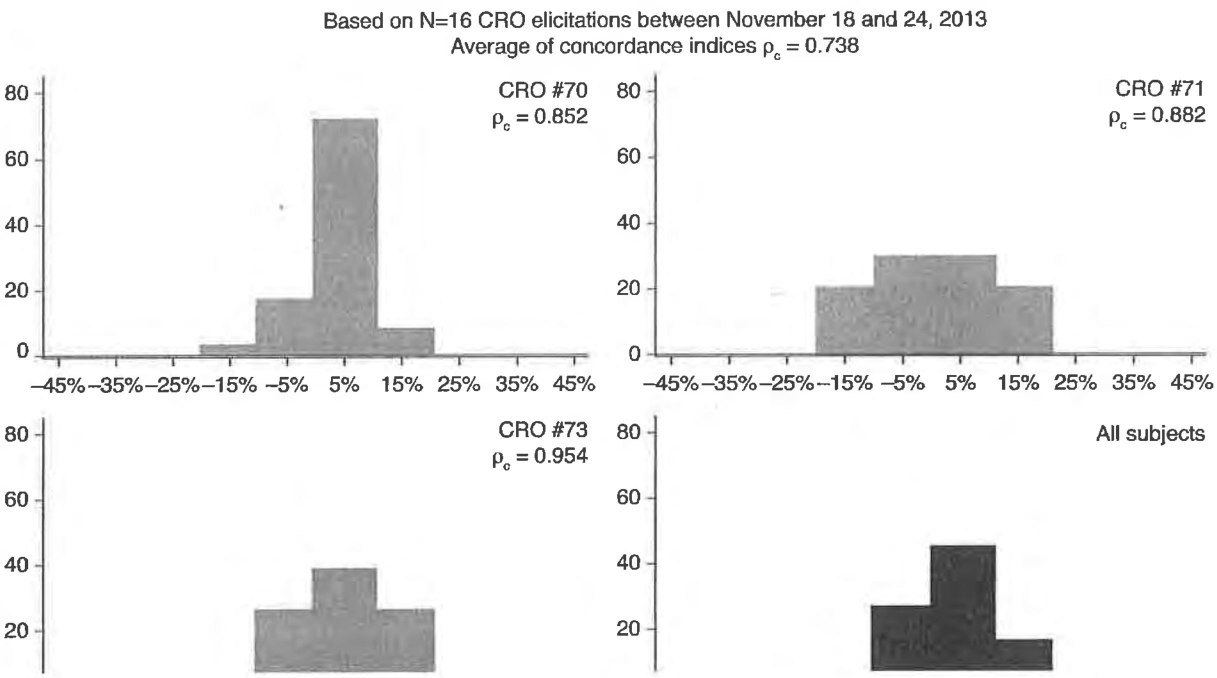


Figure 7.4 Subjective beliefs over the return on the Standard & Poors 500 Index in one year

Source:Based on N = 16 CRO elicitations between November 18 and 24, 2013; Average of concordance indices ρc = 0.738.

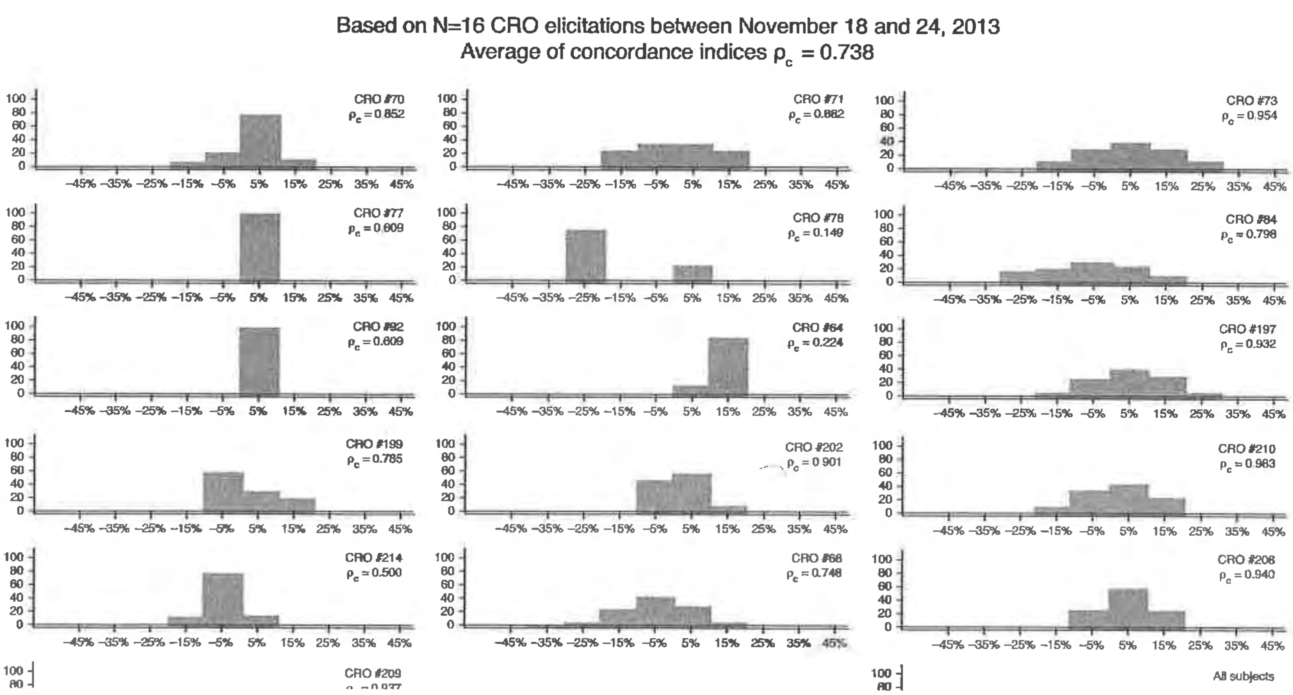


Figure 7.5 Elicited subjective beliefs of all subjects on the return on the Standard & Poors 500 Index in one year (2)

Source: Based on N = 16 CRO elicitations between November 18 and 24, 2013; Average of concordance indices ρc = 0.738.

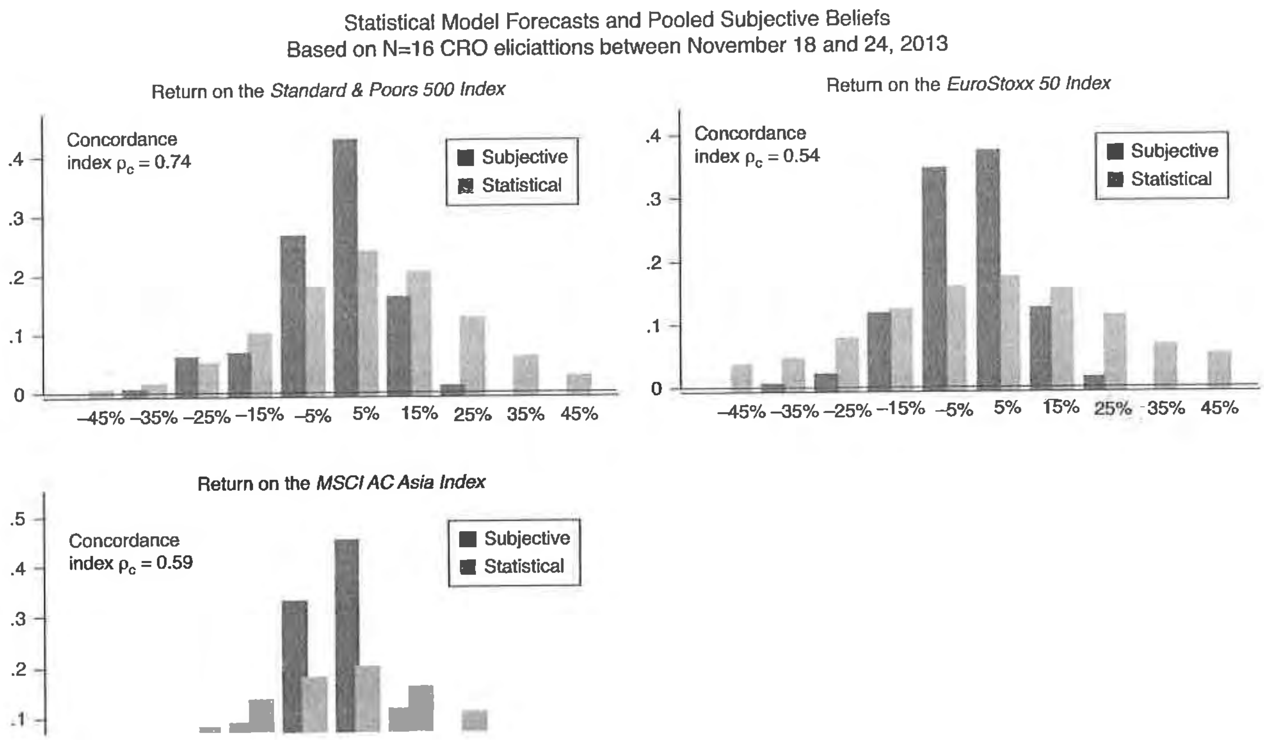


Figure 7.6 Which model of equities risk is best?

Source: Statistical Model Forecasts and Pooled Subjective Beliefs. Based on N = 16 CRO elicitations between November 18 and 24, 2013.

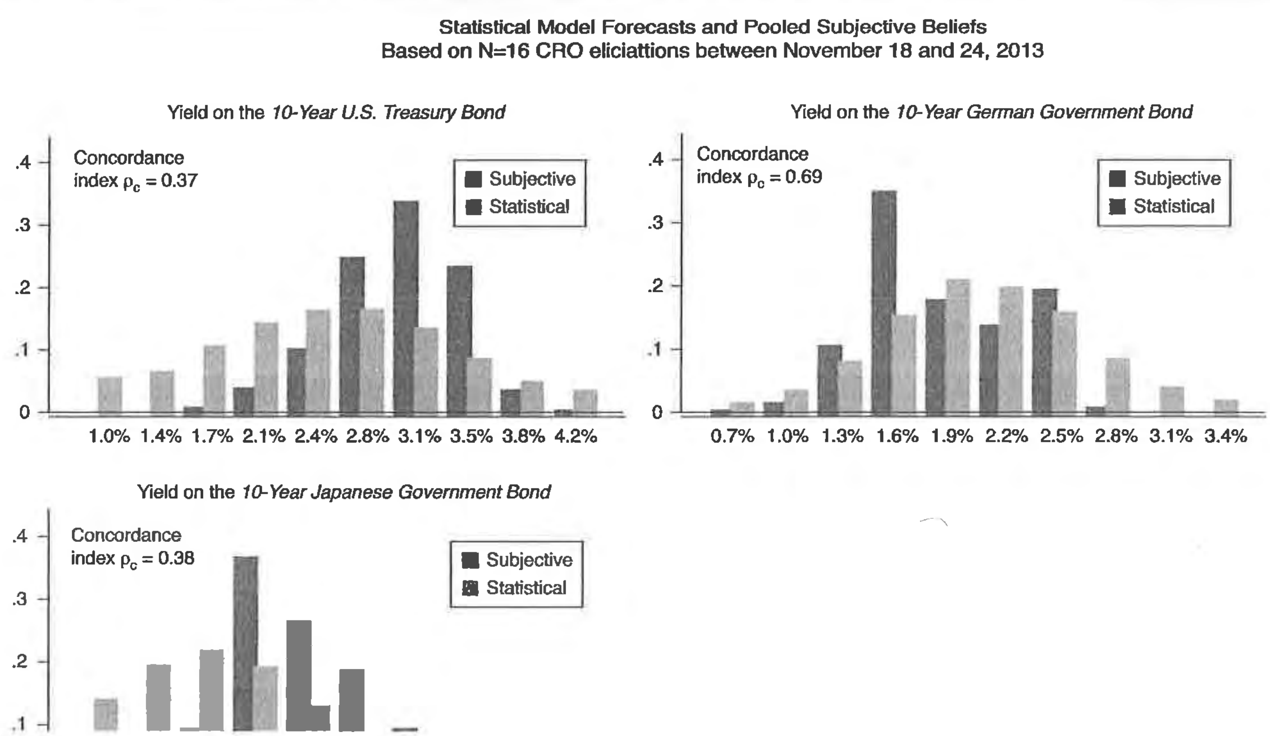


Figure 7.7 Which model of equities risk is best? (2)

Source: Statistical Model Forecasts and Pooled Subjective Beliefs. Based on N = 16 CRO elicitations between November 18 and 24, 2013.

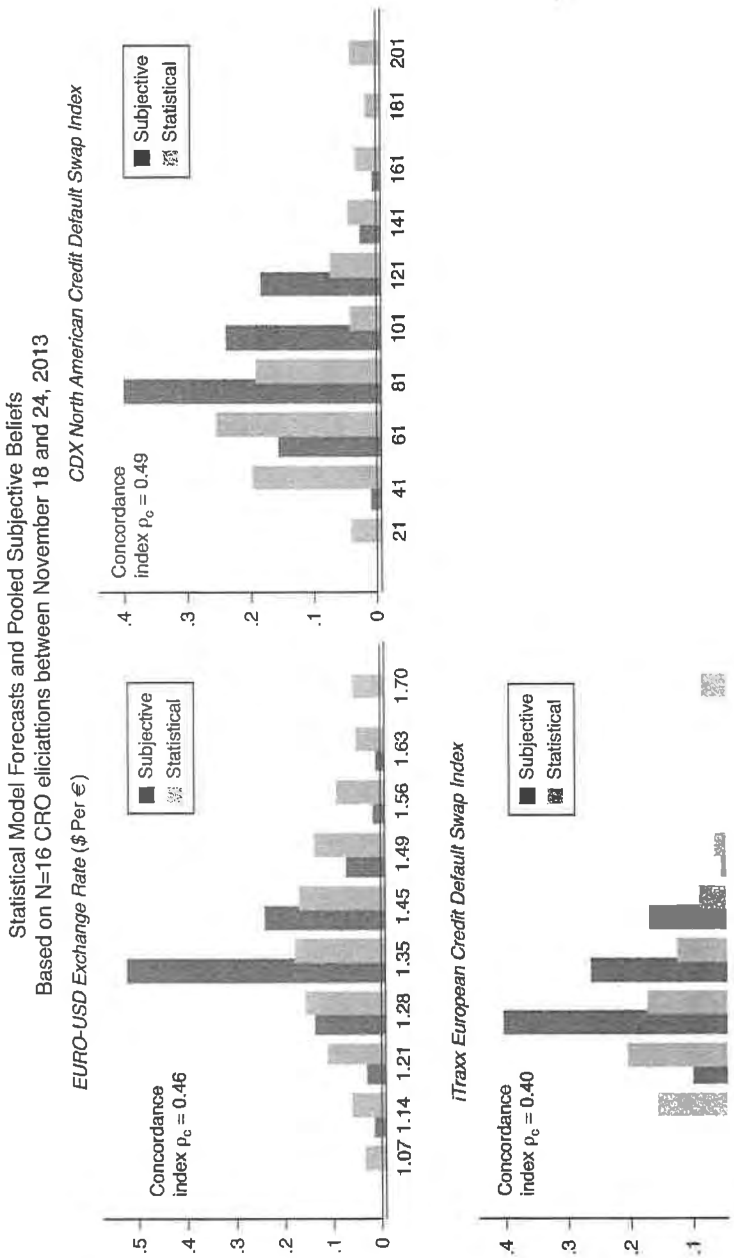


Figure 7.8 Which model of equities risk is best? (3)

Source: Statistical Model Forecasts and Pooled Subjective Beliefs. Based on N = 16 CRO elicitations between November 18 and 24, 2013.

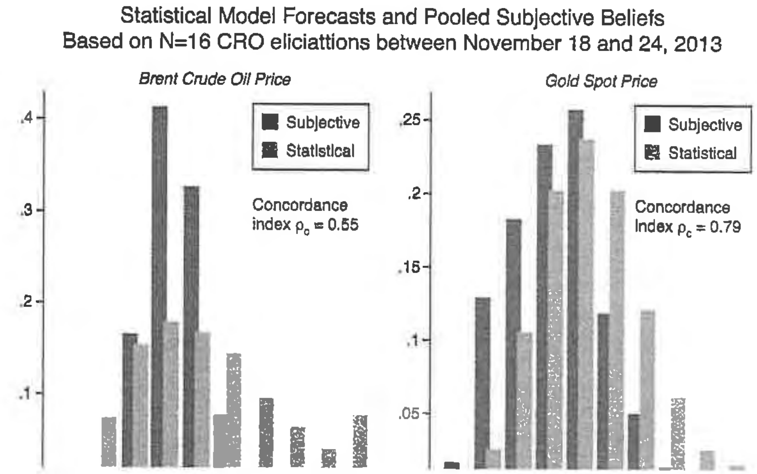


Figure 7.9 Which model of equities risk is best? (4)

Source: Statistical Model Forecasts and Pooled Subjective Beliefs. Based on N = 16 CRO elicitations between November 18 and 24, 2013.

Return" displays percentiles of the two return distributions calculated at various thresholds. The statistics under the "Probability Rate Rises" show the exceedanceprobabilities that yields on government bonds and the spreads on the CDS indices will increase above certain thresholds. The column labeled "Average CRO Concordance" displays the average of the concordance coefficients estimated between each individual subjectively elicited probability distribution relative to the pooled distribution aggregated across all CRO respondents. The number of CRO respondents was N=16.

CRO beliefs forecast modest increases in the cost of hedging investment grade credit risk over the next year. The average projection for the Markit CDX North American Investment Grade Index was 91, which falls squarely in the range of the forecast averages (88–101) observed since the February 2013 elicitation. The average subjective projection for the Markit iTraxx European Crossover Index eased to 388 from 459 in the October elicitation. Perceptions of tail risks also eased. The assessed probability of a change in the index, exceeding 100 basis points, fell to 22%, which is the lowest level observed since the first elicitation in February 2013. While forecasts on the iTraxx have been rather volatile over the course of the past year, these changes could signal an easing of concerns about Europe.

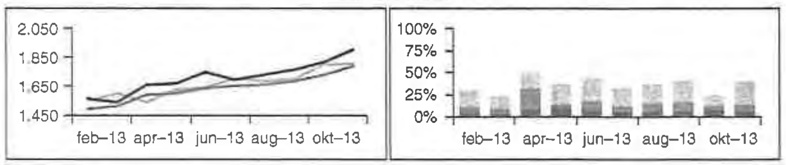
Equity Risk

Index	Statistic	Feb-13	Mar-13	Apr-13	Maj-13	Jun-13	Jul-13	Aug-13	Sep-13	Okt-13	Nov-13
S&P 500	Current Value	1.498	1.515	1.589	1.598	1.631	1.653	1.661	1.683	1.733	1.791
	1 yr Subjective Forecast	1.545	1.603	1.540	1.624	1.646	1.704	1.686	1.705	1.803	1.808
	1 yr Statistical Forecast	1.566	1.540	1.660	1.672	1.749	1.694	1.729	1.762	1.814	1.902
	< 0%	30%	23%	52%	36%	44%	33%	38%	41%	24%	40%
	< -10%	10%	9%	31%	12%	16%	12%	14%	15%	11%	13%
	< -20%	4%	2%	7%	8%	7%	5%	5%	5%	6%	6%
	Concordance	0,71	0,64	0,64	0,72	0,72	0,70	0,78	0,71	0,71	0,74
Euro Stoxx	Current Value	2.703	2.634	2.633	2.712	2.770	2.660	2.836	2.862	3.010	3.054
	1 yr Subjective Forecast	2.651	2.649	2.515	2.595	2.703	2.646	2.875	2.889	3.082	3.051
	1 yr Statistical Forecast	2.799	2.572	2.727	2.766	2.941	2.551	2.794	2.848	3.014	3.147
	< 0%	51%	44%	65%	53%	63%	46%	40%	48%	32%	49%
	< -10%	22%	17%	40%	28%	25%	22%	14%	14%	6%	14%
	< -20%	7%	4%	10%	15%	7%	6%	3%	2%	1%	2%
	Concordance	0,70	0,63	0,78	0,67	0,67	0,63	0,76	0,75	0,80	0,79
MSCI Asia	Current Value	556	555	534	552	543	501	524	536	553	533
	1 yr Subjective Forecast	546	574	515	529	549	499	500	533	571	543
	1 yr Statistical Forecast	595	572	545	571	567	475	526	538	583	551
	< 0%	57%	45%	63%	51%	54%	50%	65%	49%	37%	42%
	< -10%	34%	25%	30%	31%	28%	28%	34%	24%	20%	20%
	< -20%	11%	5%	10%	18%	11%	12%	13%	8%	8%	6%
	Concordance	0,81	0,76	0,92	0,74	0,74	0,70	0,79	0,84	0,63	0,81

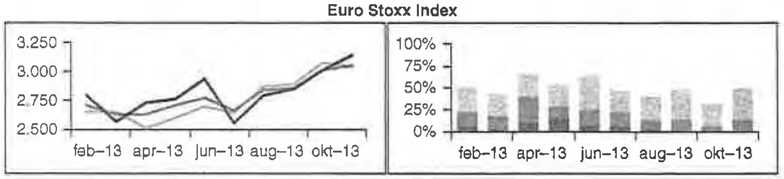
1 year Subjective Forecast
1 year Statistical Forecast
Current Value

Probability Index < 0% return (> +0bps)
Probability Index < –10% return (> +50bps)
Probability Index < –20% return (> +100bps)

S&P 500 Index



Euro Stoxx Index



MSCI Asia Index

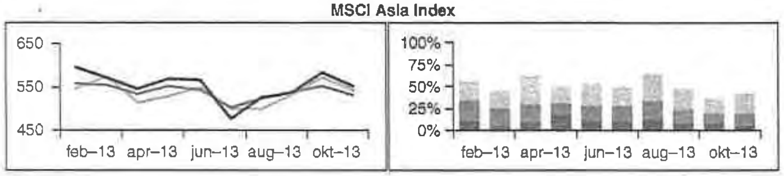


Figure 7.10 Continued

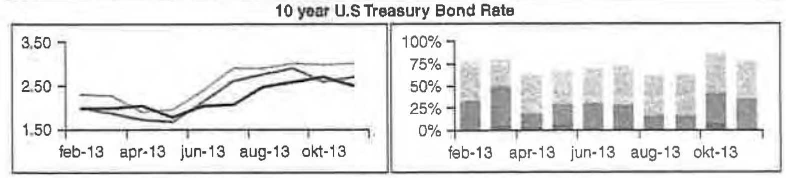
Interest Rate Risk

Index	Statistic	Feb-13	Mar-13	Apr-13	Maj-13	Jun-13	Jul-13	Aug-13	Sep-13	Okt-13	Nov-13
10 yr UST	Current Value	1,99	1,88	1,72	1,67	2,13	2,63	2,77	2,91	2,59	2,69
	1 yr Subjective Forecast	2,29	2,28	1,88	1,94	2,38	2,90	2,90	3,03	3,00	3,03
	1 yr Statistical Forecast	1,98	1,97	2,03	1,78	2,04	2,08	2,48	2,59	2,71	2,51
	> + 0bps	79%	81%	63%	67%	70%	73%	62%	64%	87%	79%
	> + 50bps	33%	49%	19%	30%	31%	29%	16%	17%	42%	37%
	> + 100bps	2%	5%	4%	7%	3%	4%	2%	3%	8%	3%
	Concordance	0,75	0,64	0,67	0,60	0,60	0,68	0,77	0,71	0,58	0,70
10 yr Bund	Current Value	1,68	1,45	1,26	1,22	1,50	1,66	1,88	2,00	1,87	1,70
	1 yr Subjective Forecast	1,77	1,57	1,33	1,46	1,70	1,86	1,99	2,13	2,13	1,83
	1 yr Statistical Forecast	1,90	1,87	1,65	1,49	1,73	1,88	1,96	2,10	2,14	2,02
	> + 0bps	54%	60%	56%	81%	71%	69%	61%	64%	75%	52%
	> + 50bps	18%	19%	10%	21%	19%	22%	13%	11%	27%	25%
	> + 100bps	1%	0%	2%	1%	3%	4%	2%	0%	5%	1%
	Concordance	0,59	0,63	0,69	0,69	0,69	0,73	0,66	0,79	0,42	0,56
10 yr JGB	Current Value	0,75	0,66	0,61	0,61	0,86	0,86	0,75	0,73	0,63	0,60
	1 yr Subjective Forecast	0,84	0,82	0,72	0,73	0,99	1,01	0,89	0,91	0,90	0,86
	1 yr Statistical Forecast	0,79	0,74	0,60	0,59	0,73	0,75	0,69	0,65	0,64	0,59
	> + 0bps	60%	73%	62%	63%	75%	76%	74%	77%	93%	90%
	> + 50bps	2%	5%	3%	7%	7%	7%	7%	6%	12%	18%
	> + 100bps	0%	0%	0%	0%	0%	0%	1%	0%	0%	0%
	Concordance	0,52	0,57	0,70	0,52	0,52	0,71	0,62	0,68	0,71	0,63

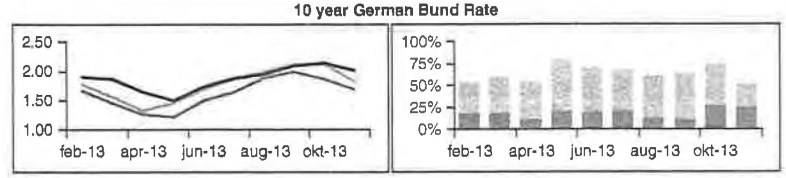
1 year Subjective Forecast
1 year Statistical Forecast
Current Value

Probability Index < 0% return (> +0bps)
Probability Index < –10% return (> +50bps)
Probability Index < –20% return (> +100bps)

10 year U.S Treasury Bond Rate



10 year German Bund Rate



10 year Japanese Government Bond Rate

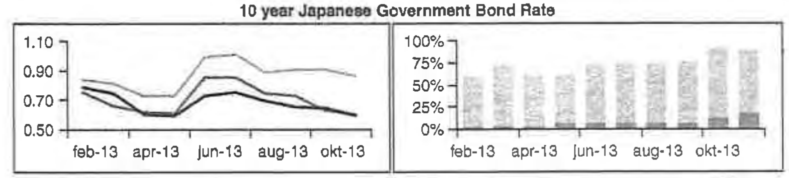


Figure 7.10 Continued

Financial Risk

Index	Statistic	Feb-13	Mar-13	Apr-13	Maj-13	Jun-13	Jul-13	Aug-13	Sep-13	Okt-13	Nov-13
Euro/USD	Current Value	1,36	1,31	1,31	1,32	1,30	1,30	1,33	1,33	1,37	1,35
	1 yr Subjective Forecast	1,32	1,33	1,32	1,28	1,28	1,31	1,35	1,33	1,38	1,36
	1 yr Statistical Forecast	1,33	1,32	1,26	1,28	1,27	1,33	1,30	1,35	1,34	1,38
	< 0%	26%	32%	52%	60%	66%	45%	38%	50%	32%	43%
	< -10%	12%	4%	6%	16%	6%	5%	4%	1%	4%	2%
	< -20%	0%	0%	0%	3%	0%	0%	0%	0%	0%	0%
	Concordance	0,68	0,80	0,77	0,74	0,74	0,81	0,70	0,84	0,76	0,79
CDX	Current Value	89	88	82	75	79	82	80	76	74	72
	1 yr Subjective Forecast	97	97	88	92	101	99	93	96	94	91
	1 yr Statistical Forecast	101	109	101	100	79	109	88	91	90	84
	> + 0bps	59%	59%	52%	75%	71%	64%	66%	75%	84%	82%
	> + 50bps	9%	5%	8%	11%	13%	17%	8%	14%	7%	11%
	> + 100bps	0%	0%	0%	0%	3%	3%	1%	2%	0%	0%
	Concordance	0,63	0,74	0,67	0,74	0,74	0,55	0,53	0,60	0,72	0,75
ITraxx	Current Value	443	447	443	396	422	441	412	387	354	351
	1 yr Subjective Forecast	401	507	511	463	510	504	494	428	459	386
	1 yr Statistical Forecast	500	558	542	554	429	586	486	484	456	408
	> + 0bps	31%	69%	73%	64%	72%	56%	73%	61%	82%	62%
	> + 50bps	22%	46%	51%	49%	58%	44%	58%	45%	59%	40%
	> + 100bps	16%	29%	31%	35%	43%	32%	42%	28%	54%	22%
	Concordance	0,64	0,85	0,94	0,70	0,70	0,68	0,82	0,69	0,65	0,76

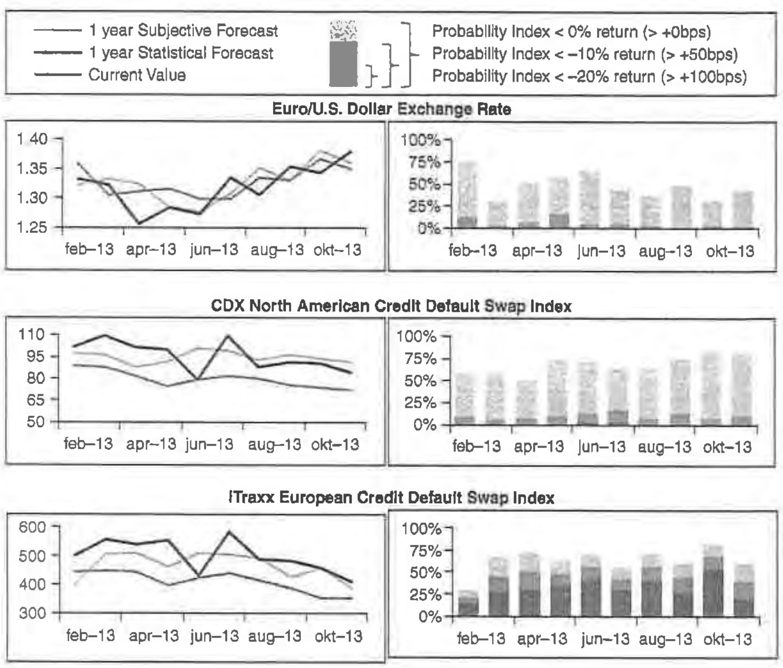


Figure 7.10 Continued

Commodity Risk

Commodity Risk

Index	Statistic	Feb-13	Mar-13	Apr-13	Maj-13	Jun-13	Jul-13	Aug-13	Sep-13	Okt-13	Nov-13
Oil	Current Value	116	111	103	102	100	109	111	113	109	109
	1 yr Subjective Forecast	126	123	112	111	110	114	115	120	121	107
	1 yr Statistical Forecast	143	144	125	118	124	118	127	133	130	129
	< 0%	31%	34%	42%	35%	36%	46%	47%	39%	38%	55%
	< -10%	11%	18%	23%	15%	13%	24%	22%	17%	18%	33%
	< -20%	3%	10%	13%	7%	6%	8%	10%	6%	3%	14%
	Concordance	0,85	0,59	0,79	0,77	0,77	0,73	0,71	0,76	0,68	0,78
Gold	Current Value	1.664	1.580	1.483	1.477	1.388	1.260	1.366	1.323	1.320	1.287
	1 yr Subjective Forecast	1.623	1.659	1.432	1.482	1.448	1.353	1.291	1.371	1.382	1.339
	1 yr Statistical Forecast	2.061	1.962	1.900	1.726	1.640	1.506	1.455	1.563	1.534	1.476
	< 0%	80%	50%	66%	65%	60%	45%	78%	59%	55%	51%
	< -10%	50%	24%	51%	34%	35%	19%	51%	28%	28%	28%
	< -20%	20%	4%	36%	13%	9%	4%	22%	4%	0%	0%
	Concordance	0,79	0,54	0,46	0,66	0,66	0,46	0,61	0,58	0,32	0,56

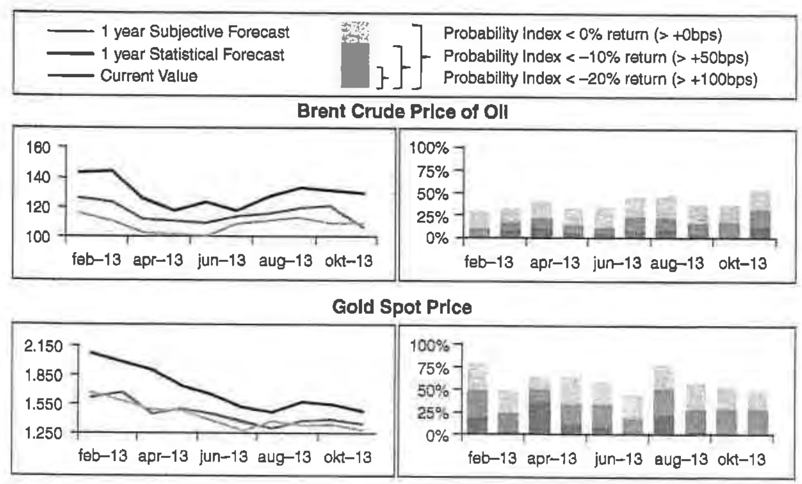


Figure 7.10 Longitudinal beliefs

The average of the CRO belief distribution eased to 1.36 from 1.38 in the September elicitation. The move coincided with a fall in the spot rate over the previous month and indicates little change in the rate from its current value. Assessed risks of large movements in the Euro/USD exchange rate continue to be very low.

Earlier in 2013 beliefs about the price of gold were pessimistic. CROs signaled that significant declines in the price of gold carried a high

Table 7.1 Summary of elicitation results for November 2013

Index	Value on 11/14/13	Forecast Index: Average	Forecast Index: Standard deviation	Percentage: Change	Percentage: Standard deviation	Probability Return: < 0%	Probability Return: < –10%	Probability Return: < –20%	Probability Rate Rises: > 0bps	Probability Rate Rises: > 50bps	Probability Rate Rises: > 100bps	Average CRO Concordance
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Standard & Poor's 500 Index												
Subjective	1,791	1,808	191	0.9%	11%	40%	13%	6%				0.74
Statistical	1,791	1,902	311	6.2%	17%	35%	17%	7%				
Eurostoxx 50 (European Blue Chip, excluding the U.K.) Index												
Subjective	3,054	3,051	298	0%	10%	49%	14%	2%				0.79
Statistical	3,054	3,147	708	3%	23%	44%	28%	16%				
MSCI AC Asia (excluding Japan) Index												
Subjective	533	543	68	2%	13%	42%	20%	6%				0.81
Statistical	533	551	165	3%	31%	45%	33%	22%				
10-Year U.S. Treasury Bond Yield												
Subjective	2.69%	3.03%	0.42%						79%	37%	3%	0.70
Statistical	2.69%	2.51%	0.82%						42%	20%	7%	
10-Year German Bund Yield												
Subjective	1.70%	1.83%	0.42%						52%	25%	1%	0.56
Statistical	1.70%	2.02%	0.56%						72%	36%	11%	
10-Year Japanese Government Bond Yield												
Subjective	0.60%	0.86%	0.23%						90%	18%	0%	0.63
Statistical	0.60%	0.59%	0.35%						45%	9%	1%	
Euro/USD Exchange Rate												
Subjective	**1.35**	**1.36**	**0.07**	**1%**	**5%**	**43%**	**2%**	**0%**				**0.79**
Statistical	**1.35**	**1.38**	**0.16**	**2%**	**12%**	**44%**	**14%**	**2%**				
CDX North American Credit Default Swap Index												
Subjective	72	91	21						82%	11%	0%	0.75
Statistical	72	84	51						50%	16%	5%	
iTraxx European Credit Default Swap Index												
Subjective	351	388	81						62%	40%	22%	0.76
Statistical	351	408	263						48%	38%	30%	
Brent Crude Oil Price												
Subjective	$109	$107	$19	–1%	17%	55%	33%	14%				0.78
Statistical	$109	$129	$54	19%	49%	40%	31%	21%				
Gold Spot Price												
Subjective	$1,287	$1,339	$186	4%	14%	51%	28%	0%				0.56
Statistical	$1,287	$1,476	$213	15%	17%	19%	5%	1%				

probability. The situation changed after significant declines in gold prices during April 2013, with the assessed risk of further significant price declines becoming small. The mean forecast is for a +4% change in gold prices over the coming year, and the assessed risk of a 20% decline in gold prices over the next year is now negligible.

Although year-ahead forecasts in the oil market have generally moved up and down with spot prices, the average forecast in the current elicitation dropped to USD 107 from 121 per barrel, despite little change in the spot price over the past month. Perceptions of downside also rose: the assessed probability of a 20% decline in oil prices over the next year now stands at 14%, vs 3% in the October elicitation.

We have been conducting the GSU CRO Risk Index since February 2013, so we now have enough months of data to produce summary information on a longitudinal basis.[19] For example, the future path of US interest rates has been the subject of much debate over 2013. Will rates rise? And, if so, will the ascent be sudden or gradual? The longitudinal results show that our CRO respondents have forecasted rising rates since the initial elicitation, and anticipate further increases from current levels. However, the path is projected to be a gradual one, since the perceived risks of sudden large movements have been and continue to be small.

In February 2013 we elicited subjective beliefs about the 11 financial risks, as at the end of December 2013. We now know the realized outcomes at the end of December 2013, and can calculate the payments to charities from the beliefs we elicited in February 2013.[20] The end-result is that four CROs earned payments that will be distributed to the following charities either anonymously or in their name: Doctors Without Borders receives $100 from two CROs, Actions Against Hunger (ACF International) receives $50, and the Humane Society of the US receives $50. Over time the earnings of respondents will be another metric for evaluating their forecasting ability.[21]

Conclusions

Modern risk management applies risk measures to forward-looking probability distributions in order to determine the amount of capital a financial institution should reserve in order to satisfy certain solvency criteria. The standard method of generating these probability distributions is typically statistical forecasts, made using familiar econometric methods for extrapolating from the past to the future. Although widely used, the methods have their limitations. For example, the models are known to be of more limited

use for long range forecasts, in markets that have a limited time series of data with which to estimate the models, and in markets where the risk manager can no longer safely assume the underlying fundamentals that generate the observed stochastic process will continue into the future.

We propose and implement an alternative methodology to generate the probability distributions needed in risk management. Drawing on insights and tools from behavioral economics, we elicit the forward-looking subjective belief distributions from Chief Risk Officers of major global financial risks using incentivized and incentive-compatible scoring rules. We compare these beliefs to the forecasts from a traditional statistical model. The extent of agreement between the subjective beliefs and statistical forecasts is formally characterized, allowing risk managers to assess for themselves what confidence to place in each. Furthermore, we characterize the extent of agreement between the individuals providing subjective beliefs, to allow an evaluation of the coherence of the beliefs that experts have about these risks.

There are many extensions of our approach. One is to accumulate a longitudinal series of elicited beliefs, and track how they change over time, correcting statistically for panel composition effects. Another is to consider weighted forecasts that utilize the self-reported expertise that each CRO has on each risk,[22] or the forecasting accuracy of the CRO as we start to have realized data that overlaps the forecasts. More challenging will be to elicit correlations of risks.

Our initial results, although based on small samples, point to some striking differences between the subjective beliefs of experts and statistical forecasts. The experts perceive significantly less upside or downside tail risk than the statistical model: in general, the standard deviation of pooled beliefs is lower. The experts had a relatively pessimistic outlook in the first half of 2013 on the risk of European and Asian equities. They expect sovereign yields in the US and Japan to rise more than the statistical model predicts, and they are generally much more optimistic about the prices of oil and gold. In these cases, either the subjective beliefs of the CRO sample or the statistical forecasters have to be closer to the true outcomes. In other instances, such as the cost of hedging credit risk in the US and exchange rate risk, the subjective beliefs of our CRO sample are broadly consistent with the forecasts of statistical models, adding strength to the inferences risk managers might draw from *those* forecasts.

Notes

Department of Risk Management & Insurance and Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University, USA. E-mail contacts: gharrison@gsu.edu and rphillips@gsu.edu. Major contributions to the research reported here were provided by Krawee Ackaramongkolrotn, Vikas Agarwal, Adam Litke, Richard Luger, Jimmy Martínez-Correa, Mark Schneider, Ajay Subramanian, J. Todd Swarthout, Eric Ulm and George Zanjani. We are grateful to the Center for Actuarial Excellence Research Fund of the Society of Actuaries for partial financial support.

1. Although we report and evaluate unweighted, pooled subjective belief distributions, since that is an interesting measure in its own right, our interest is equally in the characterization of disagreements between individual subjective belief distributions. The literature on combining evidence and expert opinion is reviewed by Shafer (1986) and Cooke (1991, Part II). There is also a related literature on how to combine density forecasts from statistical models: see Mitchell and Wallis (2011).
2. In hindsight, there are many possible precursors that provide motivation. For risk management professionals, our approach has many similarities with the concept of "credibility theory" in actuarial science (Bühlmann and Gisler, 2005). Despite its unfortunate name, which incorrectly casts doubt on the believability of one or other data source, this theory seeks to provide ways to pool information from individual risks with information from a wider class, in order to allow for a more informed judgment of risks. For instance, there might be some claim history for an individual, and yet some larger-sample statistical prior from the experience class that this individual is a member of. How does one combine these two sources? The term "credibility" comes from attaching less than 100% weight to the individual history. It is often presented as a trade-off between sampling error (low for the class, and high for the individual) and modeling error (low for the individual, and high for a heterogeneous class).
3. The literature on eliciting expert opinion in general is vast: see Cooke (1991), Garthwaite et al. (2005) and O'Hagan et al. (2006).
4. One important issue when comparing pooled (or consensus) forecasts from a panel of respondents is that changes in the composition of the panel could perfectly confound inferences about changes in the pooled forecast. Once recognized, and assuming that one has access to individual identifiers and data, there are various statistical ways to mitigate the effect of changes in panel composition: see Capistrán and Timmermann (2009) and Engelberg et al. (2011).
5. For the purposes of evaluating earnings, CROs were told that the values of all indices would be the PX_LAST value as reported on *Bloomberg* on the date stated in the interface. If that is not a business day for the index in question, we would use the closest prior date that is a business day.
6. There are also hybrids, in which responses to a small number of probability questions about the same event are used to elicit different parts of the same cumulative density function, and a distribution is then fitted to those responses. An excellent example is the evaluation of the *Survey of Economic Expectations* responses on equity returns in Dominitz and Manski (2011, §2.2).

7. For instance, see Köszegi and Rabin (2008),< Holt and Smith (2009), AQ: not present in list of references> Karni (2009) and Andersen et al. (2014).
8. The fraction of red balls in a bingo cage of red and white balls, which is briefly shown to subjects, to allow them to form a subjective belief over the true fraction.
9. More formally, we assume that utility is defined solely over the income generated by the scoring rule. If utility is event-dependent, then one must assume away any effects of the subjective outcome on initial wealth (Kadane and Winkler, 1988; Karni and Safra, 1995). In certain field applications of these scoring rules, this assumption might not be so natural. For instance, one might be eliciting beliefs about housing prices from somebody that already owns a house, so that the possible events affect the value of the initial endowment the individual has, before any income from the scoring rule. Or preferences themselves might be state-dependent, quite apart from any effect on the arguments of the utility function: different health outcomes, over which one might naturally have subjective beliefs, might affect the utility associated with given endowments. Although we believe that the incentives we offer serve to focus the attention of our CRO respondents on the elicitation task, we doubt they are integrated with their other wealth positions.
10. For instance, *Risk Talent Associates* publishes survey information on compensation packages for senior risk professionals, available at http://www.risktalent.com/cm/salary_surveys. Compensation includes salary, cash bonuses, and non-cash bonuses.
11. Appendix A of Harrison and Phillips (2013) explains the selection process. The charities were: Action Against Hunger|ACF International (http://www.acf-international.org/); American Civil Liberties Union (http://www.aclu.org/); BuildOn (http://www.buildon.org/); Doctors Without Borders (http://www.doctorswithoutborders.org/); International Federation of Red Cross and Red Crescent Societies (http://www.ifrc.org/); NAACP Legal Defense and Educational Fund (http://www.naacpldf.org/) and The Humane Society of the United States (http://www.humanesociety.org/).
12. We also encouraged the respondents to see that, by participating in a Risk Council, they were contributing to a better understanding of major financial risks. Fehr and List (2004) compare the effects of incentives for Chief Executive Officers and students in familiar laboratory experiments on "trust." They find that the former are generally more trusting and trustworthy than the latter, providing there are no threats to penalize them.
13. We use the same DJIA-generated random number x to determine which of the 11 risks to pay out for. We normalize x to the range 1 to 11 by taking the closest integer to $\frac{1}{2} + x/(100/11)$.
14. It might seem theoretically redundant to check if the "pay money" or "pay in probability" variants elicit different elicited distributions, but there is a strong behavioral rationale for undertaking this check. The reason is that the binary lottery procedure is not widely regarded by experimental economists as behaviorally reliable, whatever the theoretical claims. We view the procedure as having more behavioral validity than the received wisdom,

and critically review these concerns in Harrison et al. (2012, 2014) for objective probabilities and subjective probabilities, respectively. The empirical tests of Harrison et al. (2012) were intended to evaluate the claim that the procedure merely adds an *additional* layer of protection from the effects of non-linear utility, beyond the theoretical results suggesting that the effects are extremely weak anyway. If so, it might be attractive in applications for the binary lottery procedure to be added to practical implementations, such as the field application to CROs considered here.

15. Galati et al. (2011) conducted a *weekly* elicitation over a one-year period, starting June 22, 2010, in which respondents were asked to state what inflation rate they expected for the Netherlands in 2010, 2011 and 2012. Each question was open-ended. Subjects were staff from Dutch Central Bank, Dutch academics, and students from Dutch universities. The incentives for the elicited beliefs about 2011 and 2012 inflation were non-salient. The incentives for the elicited beliefs about 2010 employed some scoring rule, but it is not defined: "In order to obtain results accurately reflecting inflation expectations, participants were, as much as practically possible, motivated to submit their subjective beliefs by means of rewards linked to the ex-post accuracy of their expectations. This follows standard practice in the experimental economics literature." (p. 1)
16. As we conduct elicitations of subjective beliefs over time, and provide feedback to respondents on the pooled beliefs, there are also interesting hypotheses to be tested about consistency over time. Kauko and Palmroos (2014) show how one might test several such hypotheses in a Delphi forecasting context, using experts forecasting financial risks.
17. A less diplomatic statement of the same point is offered by Engelberg et al. (2009: 31ff.), who complain about the "... longstanding use of cross sectional dispersion in point predictions to measure forecaster uncertainty about future outcomes. [...] This research practice is suspect on logical grounds, even if all forecasters make their point predictions in the same way. Even in the best of circumstances, point predictions provide no information about the uncertainty that forecasters feel. This point was made forcefully 20 years ago by Zarnowitz and Lambros (1987). Nevertheless, some researchers have continued since then to use the dispersion in point predictions to measure forecaster uncertainty." Mankiw et al. (2003) and Dovern et al. (2012), for instance, use the width of the interquartile range of point forecasts as a measure of disagreement among forecasters.
18. Appropriate job titles are Chief Risk Officer ("CRO"), Chief Actuary, Vice President of Risk Management, Executive Vice President of Risk Management, Head of Credit/Interest Rate/Market Risk Management, Director of Asset Liability Management or others with a senior risk management role within the company. We were looking for senior risk professionals at international companies that are involved in a number of industries (i.e. financial institutions, conglomerates) and who have a broad view of risk across a number of markets. We limit participation to the CRO or equivalent for two reasons. First, unlike market strategists, a CRO is neutral to the outcome of results since they would not be making markets in any of the risks we cover. Second, by limiting participation to risk managers, we are soliciting the opinions of professionals who themselves are not allowed to personally participate in markets.
19. As we recruit and train additional CROs, it will be possible to use statistical techniques to control for changes in the composition of who participates in each monthly elicitation. By doing so we will be able to produce risk statistics such as those shown in the longitudinal data summary tables and charts that are independent of the mix of individual CRO respondents in each elicitation.
20. As explained in the instructions, we calculate how many "points" each CRO receives if the realized outcome is in each "bin." For instance, consider the example in Figure 7.2 for the S&P 500. If the S&P 500 actually went down by 15%, the CRO entering this forecast would have earned 28 points. If it had gone up by 8%, the CRO would have earned 88 points, and so on. We then compare the number of points earned to a random number between 1 and 100, to decide if the charity is to receive \$50 or \$0. If the points earned exceed that random number the charity earns \$50. The random number comes, as we explained, from the last two digits of the Dow Jones Industrial Average recorded at the end of the month in which the forecast ends. At the end of December 2013 this closing value was 16,576.66. Hence the last two digits are 66, implying that the points earnings would have to be 66 or more to generate \$50 for the charity. We also use this random number to decide which of the 11 risks will be used for payout, in this instance the Euro-USD exchange rate.
21. One could also compare their earnings to those of the statistical model, by "virtually placing bets" in line with the probabilistic forecast of that model each month.
22. Each CRO was asked to report how much "direct experience" they had with each of the risks. The reports were between 1 and 5, defined as follows: (1) none, (2) very little, (3) some, (4) significant, and (5) expert. We observe considerable variability, even for any one CRO.

References

Andersen S, Fountain J, Harrison GW, Rutström EE. (2014). Estimating subjective probabilities. *Journal of Risk & Uncertainty* forthcoming.

Bernanke BS, Boivin J, Eliasz P. (2005). Measuring the effects of monetary policy: A factor-augmented vector autoregressive (FAVAR) approach. *Quarterly Journal of Economics* 120(1): 387–422.

Bruine de Bruin W, Manski CF, Topa G, Van der Klaauw W. (2011). Measuring consumer uncertainty about future inflation. *Journal of Applied Econometrics* 26: 454–478.

Bühlmann H, Gisler A. (2005). *A Course in Credibility Theory*. Springer-Verlag: New York.

Capistrán C, Timmerman A. (2009). Forecast combination with entry and exit of experts. *Journal of Business & Economic Statistics* 27(4): 428–440.

Cooke RM. (1991). *Experts in Uncertainty: Opinion and Subjective Probability in Science*. Oxford University Press: New York.

Delavande A, Giné X, McKenzie D. (2011). Measuring subjective expectations in developing countries: A critical review and new evidence. *Journal of Development Economics* 94: 151–163.

Dominitz J, Manski CF. (2011). Measuring and interpreting expectations of equity returns. *Journal of Applied Econometrics* 26: 352–370.

Dovern J, Fritsche U, Slacalek J. (2012). Disagreement among forecasters in G7 countries. *Review of Economics and Statistics* 94(4): 1081–1096.

Engelberg J, Manski CF, Williams J. (2009). Comparing the point predictions and subjective probability distributions of professional forecasters. *Journal of Business & Economic Statistics* 27(1): 30–41.

Engelberg J, Manski CF, Williams J. (2011). Assessing the temporal variation of macroeconomic forecasts by a panel of changing composition. *Journal of Applied Econometrics* 26: 1059–1078.

Fehr E, List JA. (2004). The hidden cost and returns of incentives – Trust and trustworthiness among CEOs. *Journal of the European Economic Association* 2(5): 743–771.

Galati G, Poelhekke S, Zhou C (2011). Does the crisis affect inflation expectations? *International Journal of Central Banking* 7: 167–207.

Garthwaite PH, Kadane JB, O'Hagan A. (2005). Statistical methods for eliciting probability distributions. *Journal of the American Statistical Association* 100(470): 680–700.

Graham JR, Harvey CR. (2005). The long-run equity risk premium. *Finance Research Letters* 2: 185–194.

Hamilton JD. (1994). *Time Series Analysis*. Princeton University Press: Princeton, NJ.

Harrison G, Lau MI, Rutström EE. (2007). Estimating risk attitudes in Denmark: A field experiment. *Scandinavian Journal of Economics* 109(2): 341–368.

Harrison GW (2014). Hypothetical surveys or incentivized scoring rules for eliciting subjective belief distributions? Working Paper 2014–05, Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University.

Harrison GW, Martínez-Correa J, Swarthout JT. (2014). Eliciting subjective probabilities with binary lotteries. *Journal of Economic Behavior & Organization* 101: 128–140.

Harrison GW, Martínez-Correa J, Swarthout JT, Ulm E. (2012). Scoring rules for subjective probability distributions. Working Paper 2012–10, Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University.

Harrison GW, Phillips R. (2013). Subjective beliefs and statistical forecasts of financial risks: The Chief Risk Officer project. Working Paper 2013–08, Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University.

Harrison GW, Rutström EE. (2008a). Risk aversion in the laboratory, in Cox JC, Harrison GW. (eds), *Risk Aversion in Experiments*. Emerald, Research in Experimental Economics, vol. 12. Bingley, UK.

Harrison GW, Rutström EE. (2008b). Experimental evidence on the existence of hypothetical bias in value elicitation methods, in Plott CR, Smith VL. (eds), *Handbook of Experimental Economics Results*. North-Holland, Amsterdam.

Kadane JB, Winkler RL. (1988). Separating probability elicitation from utilities. *Journal of the American Statistical Association* 83(402): 357–363.

Karni E. (2009). A mechanism for eliciting probabilities. *Econometrica* 77(2): 603–606.

Karni E, Safra Z. (1995). The impossibility of experimental elicitation of subjective probabilities. *Theory and Decision* 38: 313–320.

Kauko K, Palmroos P. (2014). The Delphi method in forecasting financial markets – An experimental study. *International Journal of Forecasting* 30: 313–327.

Kim JH. (1999). Asymptotic and bootstrap prediction regions for vector autoregression. *International Journal of Forecasting* 15(4): 393–403.

Köszegi B, Rabin M. (2008). Revealed mistakes and revealed preferences. In Caplin A, Schotter A. (eds), *The Foundations of Positive and Normative Economics: A Handbook*. Oxford University Press: New York.

Lahiri K, Sheng X. (2010). Measuring forecast uncertainty by disagreement: The missing link. *Journal of Applied Econometrics* 25: 514–538.

Lichtendahl Jr. KC, Winkler RL. (2007). Probability elicitation, scoring rules, and competition among forecasters. *Management Science* 53(11): 1745–1755.

Lin LI. (1989). A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 45: 255–268.

Lin LI. (2000). A note on the concordance correlation coefficient. *Biometrics* 56: 324–325.

Mankiw NG, Reis R, Wolfers J. (2003). Disagreement on inflation expectations, in Gertler M, Rogoff K. (eds), *NBER Macroeconomics Annual 2002*. MIT Press: Cambridge, MA.

Manski CF. (2004). Measuring expectations. *Econometrica* 72(5): 1329–1376.

Matheson JE, Winkler RL. (1976). Scoring rules for continuous probability distributions. *Management Science* 22(10): 1087–1096.

Mitchell J, Wallis KF. (2011). Evaluating density forecasts: Forecast combinations, model mixtures, calibration and sharpness. *Journal of Applied Econometrics* 26: 1023–1040.

Nickerson CAE. (1997). A note on "A concordance correlation coefficient to evaluate reproducibility." *Biometrics* 53: 1503–1507.

O'Hagan A, Buck CE, Daneshkhah A, Eiser JR, Garthwaite PH, Jenkinson DJ, Oakley JE, Rakow T. (2006). *Uncertain Judgements: Eliciting Experts' Probabilities*. Wiley: Chichester, UK.

Perozek M. (2008). Using subjective expectations to forecast longevity: Do survey respondents know something we don't know? *Demography* 45(1): 95–113.

Savage LJ. (1971). Elicitation of personal probabilities and expectations. *Journal of American Statistical Association* 66: 783–801.

Savage LJ. (1972). *The Foundations of Statistics*, 2nd ed. Dover Publications: New York.

Shafer G. (1986). The combination of evidence. *International Journal of Intelligent Systems* 1: 155–179.

Sims CA. (1980). Macroeconomics and reality. *Econometrica* 48: 1–48.

Smith CAB. (1961). Consistency in statistical inference and decision. *Journal of the Royal Statistical Society* 23: 1–25.

Thombs LA, Schucany WR. (1990). Bootstrap prediction intervals for autoregression. *Journal of the American Statistical Association* 85: 486–492.

Vissing-Jorgensen A. (2004). Perspectives on behavioral finance: Does 'irrationality' disappear with wealth? Evidence from expectations and actions, in Gertler M, Rogoff K. (eds), *NBER Macroeconomics Annual 2003*. MIT Press: Boston, MA.
Winkler RL, Murphy AH. (1970). Nonlinear utility and the probability score. *Journal of Applied Meteorology* 9: 143–148.
Zarnowitz V, Lambros LA. (1987). Consensus and uncertainty in economic prediction. *Journal of Political Economy* 95(3): 591–621.

8

Defaults and Returns in the High-Yield Bond and Distressed Debt Market: Review and Outlook

Edward I. Altman and Brenda J. Kuehne

Defaults, default rates and recoveries

High-yield bond default rates on US, Canadian and Mexican high-yield bonds decreased slightly in 2013 and remained well below historical averages. The rate decreased from 1.62% at year-end 2012 to 1.04% for all of 2013. Defaults include straight corporate bonds whose firms went bankrupt, missed an interest payment and did not cure it within the grace or forbearance period, or completed a distressed exchange. The 2013 rate is based on a mid-year market size of $1.39 trillion, up by a sizeable $180 billion from a year earlier. In all, $14.5 billion of defaults were recorded in 2013 (Table 8.1). The historical weighted-average annual default rate is 3.61% over the 43-year period (1971–2013). This weighted-average rate is down compared to 3.82% at the end of 2012. Our weights are based on the par value of high-yield bonds outstanding in each year. The arithmetic annual average default rate dropped to 3.14% from 3.19% one year earlier.

The fourth-quarter 2013 default rate was 0.18%, lower than one year earlier (0.57%), and the lowest quarterly default rate since the second-quarter 2011. This continues a trend of low defaults, during which the quarterly default rates were below 0.50% in fourteen out of the last sixteen quarters (1Q 2010–4Q 2013), with only the fourth-quarters of 2011 and 2012 default rates higher than this level. During the period, defaults remained below 0.50% for seven consecutive quarters – from 1Q 2010 to 3Q 2011. Since 1989 there have been two longer, consecutive quarterly periods of default rates also below 0.5% – seven from 4Q 2003 to 2Q 2005 and nine from 1Q 2006 to 1Q 2008 (Figure 8.1).

Contemporary Challenges in Risk Management

Dealing with Risk, Uncertainty and the Unknown

Edited by

Torben Juul Andersen
Copenhagen Business School, Denmark

palgrave
macmillan

First published 2014 by
PALGRAVE MACMILLAN

Palgrave Macmillan in the UK is an imprint of Macmillan Publishers Limited, registered in England, company number 785998, of Houndmills, Basingstoke, Hampshire RG21 6XS.

Palgrave Macmillan in the US is a division of St Martin's Press LLC, 175 Fifth Avenue, New York, NY 10010.

Palgrave Macmillan is the global academic imprint of the above companies and has companies and representatives throughout the world.

Palgrave® and Macmillan® are registered trademarks in the United States, the United Kingdom, Europe and other countries

ISBN: 978–1–137–44760–9

This book is printed on paper suitable for recycling and made from fully managed and sustained forest sources. Logging, pulping and manufacturing processes are expected to conform to the environmental regulations of the country of origin.

A catalogue record for this book is available from the British Library.

A catalog record for this book is available from the Library of Congress.

Contents