RESOURCES *for the* FUTURE

The Social Cost of Carbon: Advances in Long-Term Probabilistic Projections of Population, GDP, Emissions, and Discount Rates

Online Appendix

Kevin Rennert, Brian C. Prest, William A. Pizer, Richard G. Newell, David Anthoff, Cora Kingdon, Lisa Rennels, Roger Cooke, Adrian E. Raftery, Hana Ševčíková, and Frank Errickson

October 2021

Contents

I. Description of Censoring of Raw MSW Economic Growth Dataset

As indicated in MSW (2019), there is considerable uncertainty in future economic growth 100-300 years into the future. As a result, the tails of the MSW distribution are quite wide, leading to some implausibly small or implausibly large future levels of GDP per capita in the extreme tails (e.g., below the 1st percentile or above the 99th percentile). These extreme tails correspond to extremes of persistent economic growth beyond that which has been observed for any country in the historical record over such long time periods (e.g., below -1% or above +5% annually on average through 2300), but nonetheless are possible as simulated projections given the distributional assumptions of the MSW model. Such low or high sustained growth rates would lead to global GDP/capita either falling by more than 90% between 2021 and 2300 (e.g., 0.99^{279}) or rising by a factor of more than 800,000 (1.05^{279}) implying a global average income of more than $10 billion per person. However, according to the Maddison Project dataset,[1] which includes country-level GDP/capita data as far back as 1500 for some countries, no country has ever experienced such extreme growth for such long periods of time.[2] Further, the 1st and 99th percentiles of the combined distribution of long-run growth rates based on the EGS are -0.6% and +4.4%, indicating long-run growth rates are exceptionally unlikely to fall outside this range. In the MSW model, those extreme tail outcomes are very likely driven by the structure of the Bayesian model, such as its embedded distributional assumptions, rather than the historical data used to estimate the model.

For these reasons, and in consultation with James Stock (one of the authors of MSW (2019)), we lightly censor some projections in the extreme tails of the MSW distribution that are outside the range of historical experience and also outside the long-run range implied by the EGS, using a combination of two approaches. We first use the Maddison Project data to calculate, for each annual time horizon from 1 to 150 years, the minimum and maximum sustained growth rates ever observed by any country historically over that time horizon. For example, at the 50-year time horizon, these minimum and maximum rates were -2.4% (DR Congo, 1954-2004) and +7.0% (Equatorial Guinea 1958-2008). While this approach is sufficient to establish bounds for horizons up to 150 years into the future, it is less informative for the very long term (e.g., 150-300 years) because the Maddison data has relatively few countries for such long time horizons. Therefore, we augment these historical ranges using the range suggested by the extreme quantiles of the combined growth distribution from the EGS. Specifically, we use as bounding values the 0.5th and 99.5th quantiles of the combined (performance-weighted) distribution of long-run (2020-2300) growth rates from the EGS, which are -0.88% and +5.1%.

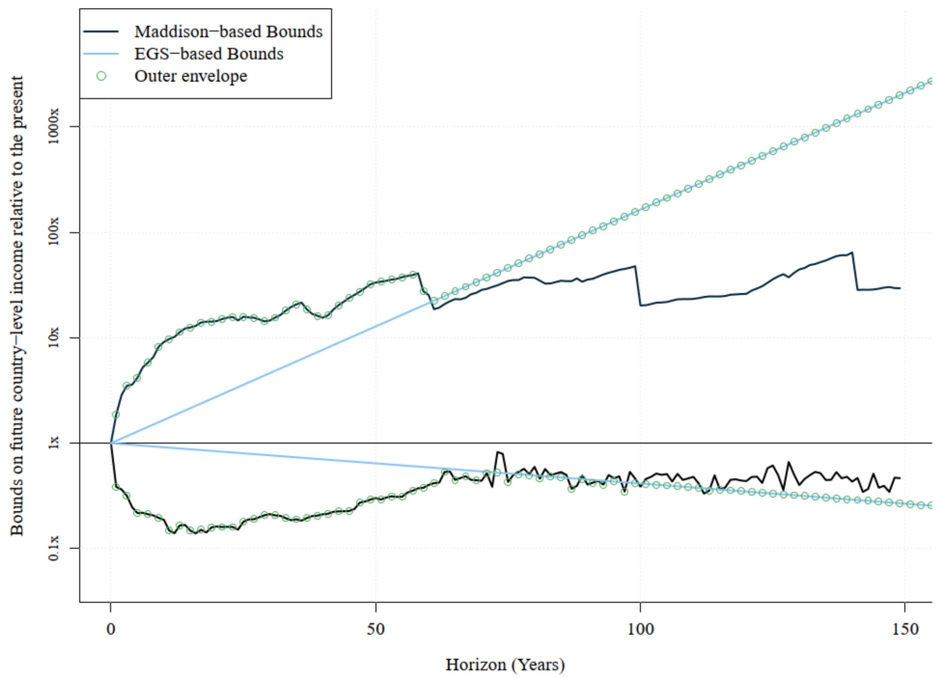
---

[1] Available at **https://clio-infra.eu/Indicators/GDPperCapita.html**

[2] For example, no country in Maddison Project data has observed 100-year growth rates of below -1% or above +3%,

We use the outer envelope of (i.e., the less restrictive of) these two sets of bounding values for growth for each time horizon from 1 to 300 years. Specifically, these bounds are calculated as the outer envelope of the change in log(GDP/capita) implied by the horizon-specific growth rates from the Maddison data and those implied by the EGS (-0.88% and +5.1% annually). Country-by-year values in the projections that exceed these absorbing bounds are set equal to the bounds. This affects less than one percent of country-by-year draws in each tail of the distribution. The time paths of these absorbing bounds are shown in Figure OA-1.

Figure OA-1.Absorbing Bounds for Change in Country-Level GDP/capita Over Time.



II. Structured Expert Judgment: Details on experts and scoring for Economic Growth Survey (EGS) and Future Emissions Survey (FES)

1. Introduction

This report details the structured expert judgment studies of GDP and greenhouse gas emissions conducted by RFF as part of RFF's activities to implement recommendations from the National Academies of Sciences, Engineering, and Medicine (NASEM 2017) for improving estimates of the social cost of carbon.

Ten experts participated in the Economic Growth Survey (EGS) in Washington DC in 2019-2020.[3] Ten experts participated in the Future Emissions Survey (FES) elicitation also held at RFF in 2021. From previous experience a working minimum panel size is about six experts and more than 20 provides diminishing returns in terms of the performance of the pooled expert judgments.

The participating experts in the EGS were: Daron Acemoglu, Erik Brynjolfsson, Jean Chateau, Melissa Dell, Robert Gordon, Mun Ho, Chad Jones, Pietro Peretto, Lant Pritchett, Dominique van der Mensbrugghe. The participating experts in the FES were: Sally Benson, Geoff Blanford, Leon Clarke, Elmar Kriegler, Jennifer Faye Morris, Sergey Paltsev, Keywan Riahi, Susan Tierney, Detlef van Vuuren. Elicitations were conducted face to face remotely. Experts were provided with an introductory video and background materials to orient them to the study.

The expert judgment methodology applied here is termed the "*Classical Model*" because of its analogy to classical hypothesis testing (1). The key idea is that experts are treated as statistical hypotheses. They are scored as uncertainty assessors based

[3] One expert opted out of the calibration process. Results from that expert were therefore included only in the equal weight combinations discussed below.

on their responses to calibration variables from their field whose true values are unknown to the experts at the time of the elicitation. The purpose of scoring is twofold. First scoring enables performance weighted combinations of experts' judgments. Second, the scores of combinations of experts serve to gauge and hopefully validate the combination which is adopted.

2. Expert Scoring

Each expert stated 5th , 50th and 95th percentiles, or quantiles, for each calibration variable. An expert's *statistical accuracy* is the P-value (column 2 in Tables OA-3 and OA-4) at which we would falsely reject the hypothesis that an expert's probability assessments are statistically accurate. Roughly, an expert is statistically accurate if, in a statistical sense, 5% of the realizations fall beneath the expert's 5th percentile, 45% of the realizations fall between the 5th and 50th percentile, etc. High values (near 1) are good, low values (near 0) reflect low statistical accuracy. An expert's *informativeness* is measured as the mean Shannon relative information in the expert's distribution with respect to a uniform background measure over an interval containing all experts' percentile assessments and the realizations, variable-wise[4].

Column 3 of Table OA-1 gives the average information scores for each expert for all calibration variables. The number of calibration variables is shown in column 4 for each expert (in this case all experts assessed all 11 calibration variables). The product of columns 2 and 3 is the *combined score* for each expert (not shown). Note that statistical accuracy scores vary over four orders of magnitude whereas information scores vary within a factor two. Statistical accuracy is a fast function while informativeness is slow. Therefore, by design, the ratios of the products of combined scores are dominated by the statistical accuracy.

There is a loose negative correlation between experts' statistical accuracy and information. Note also that while the equal weight combination's statistical accuracy in both panels is comfortably above the 5% threshold, its information is lower than that of any expert. As mentioned, information is a slow function; halving the information score corresponds roughly to doubling the size of the 90 percent confidence bands. This information penalty of equal weighting is typical of other expert judgment panels.

Six of the nine EGS experts had a statistical accuracy score above 0.05 which is the traditional threshold below which statistical hypotheses would be rejected. For the FES panel this holds for 5 of the 10 experts. A recent review of the 49 studies

---

[4] Computing the mean relative information requires fitting densities to each experts' quantile assessments. The minimal informative density relative to the background measure which complies with the expert's quantiles is chosen for this purpose. The mean relative information is proportional to the information in each expert's joint distribution if the distributions for each variable are independent. The mean is taken to prevent the importance of informativeness to depend on the number of calibration variables. The mean relative information is a global performance measure; the actual weights employ the information scores per variable and are thus variable specific.

conducted since 2006 found that 75% of the 530 experts would be rejected as statistical hypotheses at the 5% level (6). In this sense, both the EGS and the FES panels display an unusually high number of non-rejected experts. This in turn leads to lower performance differences between equal and performance weighting than is seen in many other studies.

Table OA-1. Scores and weights for all 9 EGS experts when performance weights are not optimized but computed for the six weighted experts with statistical accuracy > 0.05.

		Expert Scores EGS			
expert	**Statistical accuracy**	**mean information**	**# variables**	**weight**	**Rel.Inf to EW DM**
1	0.706	0.673	11	0.3	0.433
2	0.399	0.829	11	0.209	0.608
3	0.008	0.894	11	-	0.6
4	0.197	0.659	11	0.082	0.363
5	0.215	1.094	11	0.148	0.711
6	0.327	1.131	11	0.233	0.761
7	0.154	0.291	11	0.028	0.331
8	0.018	1.087	11	-	0.671
9	0.0003	0.727	11	-	0.511
PW05	0.492	0.457	11	-	
EW	0.37	0.266	11	-	

Notes: Statistical accuracy denotes the significance level at which the hypothesis that an expert is statistically accurate would be falsely rejected. Mean Information denotes the average Shannon relative information in an expert's assessments for all calibration variables. "# variables" denotes the number of calibration variables answered by an expert. "Weight" for weighted experts is the normalized sum of the product of columns 2 and 3. "Rel Inf to EW DM" is an expert's relative information with respect to the EW combination of all experts.

Table OA-2. Scores and weights for all 10 FES experts when performance weights are optimized (Op) and also not optimized (nOP, no statistical accuracy cut-off is applied).

Expert Scores FES					
expert	**Statistical accuracy**	**mean information**	**# variables**	**weight (GWnOp)**	**Rel.Inf to EW DM**
1	0.083	0.893	11	0.079	0.636
2	1.26E-05	1.354	11	0.000	1.003
3	1.70E-07	1.592	11	0.000	1.182
4	3.01E-04	1.357	11	0.000	1.271
5	0.003	1.295	11	0.005	0.935
6	0.399	0.548	11	0.234	0.400
7	0.018	0.748	11	0.015	0.649
8	0.169	0.676	11	0.122	0.460
9	0.327	0.846	11	0.295	0.601
10	0.385	0.611	11	0.251	0.408
GWnOp	0.615	0.308	11		0.006
GWOp	0.852	0.313	11		0.043
IWnOp	0.385	0.361	11		0.042
IWOp	0.615	0.361	11		0.210
EW	0.638	0.238	11		0

Notes: The weights in column 5 are those of GWnOP. Global weighting (GW) uses the mean informativeness over all calibration variables. Statistical accuracy denotes the significance level at which the hypothesis that an expert is statistically accurate would be falsely rejected. Mean Information denotes the average Shannon relative information in an expert's assessments for all calibration variables. "# variables" denotes the number of calibration variables answered by an expert. "Global Weight" for weighted experts is the normalized sum of the product of columns 2 and 3. Item specific weighting (IW) applies different weights for each item based on an experts' informativeness per item. "Rel Inf to EW DM" is an expert's relative information with respect to the EW combination of all experts.

3. Combining Experts

A scoring system is asymptotically strictly proper if an expert obtains the expert's highest expected score in the long run by, and only by, stating percentiles corresponding to the expert's true beliefs. Statistical accuracy and informativeness are dimensionless. Their product, termed the combined score, is an asymptotically strictly proper scoring rule if experts get zero weight when their P-value drops below some positive threshold (1). If the expert tries to game the system to maximize the expert's expected weight, the expert will eventually figure out that honesty is the only optimal strategy. The theory does not say what the cut-off value should be, so this is chosen on the basis of extra-mathematical considerations. One strategy is to choose the cut-off at the level which optimizes the combined score of the resulting decision maker. Another strategy is to choose a statistical cutoff, typically 0.05, such that weighted experts are those who would not be rejected as statistical hypotheses. A third strategy is to choose a cutoff sufficiently low that all empanelled experts are weighted (termed not optimized or nOp).

For the EGS, optimal weighting resulted in expert 1 receiving weight one, other experts being unweighted. The strategy of choosing the cutoff at the traditional 5% value, termed PW05 shown in Table OA-2. All three DM's EW, PWOp and PW05 show acceptable statistical performance[5]. EW exhibits lower informativeness than any expert and lower than PWOp and PW05. Very roughly, halving the informativeness corresponds to doubling the width of the 90% confidence intervals. In this case PWOpt was rather non robust whereas PW05 was very robust (see below).

For the FES there was very little difference in performance between IWOp, IWnOp, GWOp and GWnOp. All are more informative than EW. Optimization incurs a penalty in robustness (see below) and in light of the small differences in performances, preference was given for the non-optimized versions.

The Classical Model has been applied in hundreds of expert panels and has been validated both in- and out-of-sample (2-6). In the absence of observations of the variables of interest, out-of-sample validation comes down to cross-validation whereby the calibration variables are repeatedly separated into subsets of training- and test variables. The PW model is initialized on the training variables and scored on the test variables. The superiority of PW over EW in terms of statistical accuracy and informativeness has been demonstrated using this approach (2).

[5] If an expert is statistically accurate, then his/her statistical accuracy score is uniformly distributed on the interval [0, 1] and small values are significant. A difference between 0.4 and 0.5 is not important, but a difference of 0.4 and 0.008 is.

4. Robustness

Robustness on experts examines the effect on the preferred combination of losing individual experts. For the EGS panel, experts are removed one at a time and PW05 is recomputed as shown in Table OA-3.

Table OA-3. Robustness on EGS experts

Robustness analysis on Experts			
Excluded expert	**Information wrt background**	**Statistical accuracy**	**Information wrt original DM**
1	0.524	0.492	0.106
2	0.511	0.370	0.105
3	0.428	0.492	0.007
4	0.478	0.492	0.057
5	0.445	0.492	0.037
6	0.424	0.492	0.054
7	0.391	0.492	0.043
8	0.451	0.492	0.003
9	0.393	0.492	0.011

Notes: For explanations of columns see section 1.2, "Information wrt original DM" gives the mean relative information of the perturbed DM (with one expert removed) with respect to the original DM.

Table 4 shows that the mean relative information with respect to the original PW05 is in the order 0.05. Comparison with the values in the rightmost column of Table 2 shows that the changes wrought in PW05 by loss of a single expert are much smaller than the differences among the experts themselves. In this sense we conclude that PW05 is robust against the loss of any single expert. Had we chosen PWOpt, the difference caused by losing the expert with weight 1 would be 0.56, which is on the order of the expert differences.

Robustness against loss of a calibration variable proceeds in the same manner. Calibration variables are removed one at a time and PW05 is recomputed. These scores are extremely robust against loss of a calibration variable. A calibration variable may exert leverage on the combination of experts if all experts assess this variable poorly. That may be due to "group think" or it may be due to inappropriateness, ambiguity or error in the true value. If one variable perturbs the combination more than the others, this flags the variable in question for further scrutiny. Table 5 raises no flags in this regard. Comparing the rightmost columns of Tables 1 and 5 show that the perturbation caused by loss of a single calibration variable is small relative to the divergence among the experts themselves.

Table OA-4. Robustness on EGS calibration questions (calib vbl)

	Robustness analysis on variables		
Excluded variable	**Information wrt background**	**Statistical accuracy**	**Information wrt original DM**
GrwCh	0.44	0.47	0.04
GrwSA	0.49	0.47	0.04
GrwSSA	0.44	0.47	0.05
MADCh	0.41	0.55	0.06
MADSA	0.51	0.55	0.05
MADSSA	0.52	0.47	0.05
CBOErr	0.45	0.55	0.04
$StLIC	0.38	0.47	0.04
$StHIC	0.45	0.55	0.02
DyCLIC	0.47	0.55	0.04
DyCHIC	0.50	0.47	0.05

Notes: For explanations of columns see section 1.2. "Information wrt original DM" gives the mean relative information of the perturbed DM (with one expert removed) with respect to the original DM.

Table OA-5. Robustness on FES experts and items for optimized vs non optimized global weights

excluded expert	Rel Inf wrt PWnOp	rel inf wrt GWOp	Excluded item	Rel Inf wrt PWnOp	rel inf wrt GWOp
1	0.01695	0.2578	spAdEc	0.0131	0.04904
2	2.51E-06	3.39E-09	spEmMr	0.0231	0.298
3	1.79E-07	0	spEaPc	0.02474	0.2561
4	0.006481	0.2578	%OECD	0.0144	0.3188
5	0.001533	2.67E-08	%SSAf	0.07927	0.04931
6	0.08502	0.1604	#ngSA	0.07836	0.3087
7	0.009123	0.06184	#ngSSA	0.01785	0.3204
8	0.0172	0.03377	%ergME	0.06166	0.4397
9	0.04194	0.2578	%ergCh	0.02913	0.2442
10	0.07871	0.08283	RnwVen	0.05453	0.06587
			RwnPor	0.02095	0.3209

5. Random Expert Hypothesis

The random expert hypothesis states that putative differences in performance between assessors is just noise and does not indicate persistent differences of the assessors (6). One way to test this hypothesis is to compare panel wide performance metrics in the original panel with the same metrics as generated by a large set (here 1000) of "scrambled panels" in which the assessments are randomly re-allocated to assessors, thus wiping out any 'assessor effect'. Considering statistical accuracy (SA) and information (Inf), we are interested in the panel maxima, minima averages and standard deviations. Table 6 shows, for example, that the average SA score in the original EGS panel was 0.23. In 14.1% of the 1000 scrambled panels the average SA score was lower than 0.23. The minimum SA score in the original EGS panel was 0.0003 and in all 1000 scrambled panels the minimum SA score was greater than the original panel minimum Sa. Although the scrambling was able to exceed the panel average SA in 85.9% of the cases, it was never able to get scores as low as the minimum in the original panel. The average Inf score is always the same in the scrambled and original panels, but the scrambling was unable to reproduce the highest and lowest scores. The same pattern is observed in the second panel. The fact that significant departures from randomness (indicated by italicized values) do not occur for average or maximum values of statistical accuracy may reflect the fact that there relatively large numbers of statistically accurate experts in each panel. The random scrambling is unable to reproduce the extremes of the information scores. The conclusion that differences in expert performance arise by chance is not supported in either panel.

Table OA-6. Results of testing the random expert hypothesis in the EGS and FES panels

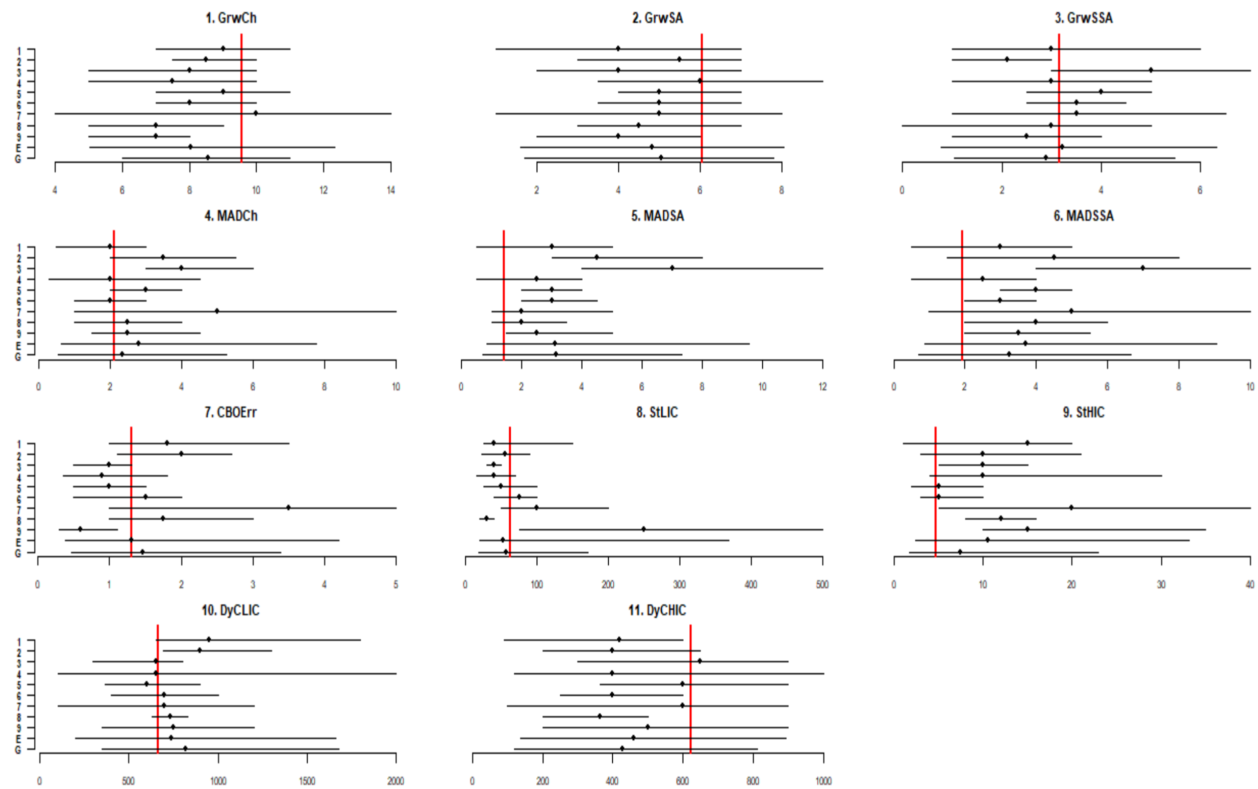
		Panel metrics				In Random re-allocations			
		average	**StDev**	**max**	**min**	**%<ave**	**%<stdev**	**%<max**	**%>min**
EGS	**SA**	0.23	0.23	0.71	3.00E-04	14.10%	30.5%	35.9%	91.2%
	inf	0.83	0.27	1.14	0.30	-	*100%*	95.7%	*100%*
FES	**SA**	0.14	0.17	0.40	1.70E-07	93.30%	76.5%	55.3%	*98.0%*
	inf	1.00	0.37	1.60	0.55	-	*100%*	*99.7%*	*98.0%*

Notes: Panel wide metrics for statistical accuracy (SA) and informativeness (Inf) in the original panel are compared with1000 values in each of 1000 random re-allocations of assessments. The random re-allocation wipes out any "assessor effect", thus values greater than 95% in the rightmost 4 columns indicate significant departures from randomness.

6. Range Graphs

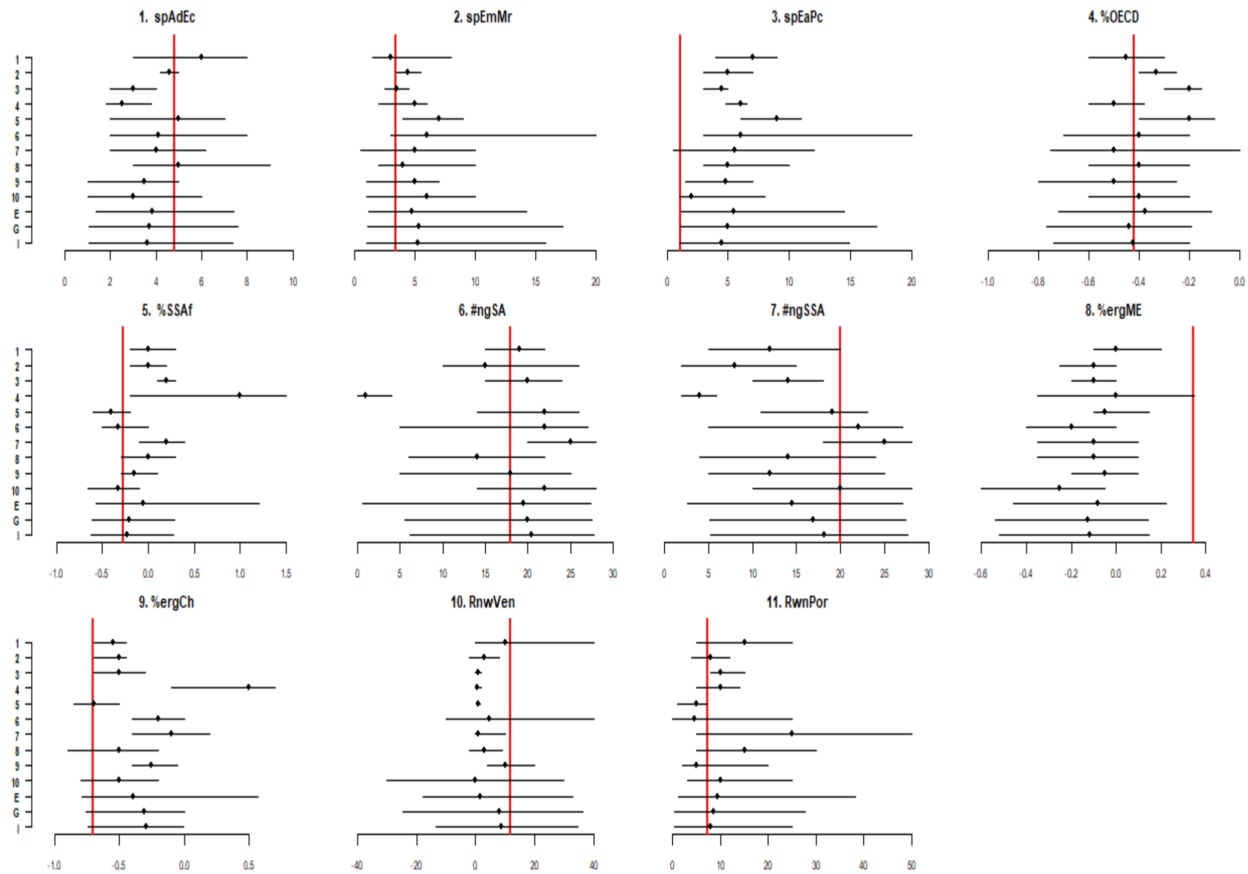
Range graphs give a graphic representation of all assessments of all variables together with the true values.

Figure OA-2. Range graphs for calibration variables for the EGS: one stacked graph for each variable



Notes: Each assessment is given by a horizontal line with 5% and 95% values as endpoints and a dot for the median. The true value is indicated by a vertical line. On each graph the order of assessments from top to bottom is expert 1, expert 2, …expert 9, EW, PW05.

Figure OA-3. Range graphs for calibration variables for the FES: one stacked graph for each variable



Notes: Each assessment is given by a horizontal line with 5% and 95% values as endpoints and a dot for the median. The true value is indicated by a vertical line. On each graph the order of assessments from top to bottom is expert 1, expert 2, ...expert 10, EW, GWnOp, IWnOp.

Each assessment is represented as a horizontal line whose endpoints correspond to the 5th and 95th percentiles with a dot representing the median. The true value is shown as a red vertical line. Range graphs are helpful in identifying any variables for which experts show structural differences: some experts may be too low while others are too high, some experts may be very confident (narrow bands) while others are very uncertain (wide bands), some experts may be isolated (non-intersecting confidence bands). Relative to other expert panels, the graphs in Figures OA-2 and OA-3 are very coherent. There are no isolated experts, and no evidence of diverging schools of thought.

7. Conclusion

The expert data presented here is unusual in the number of experts with high statistical accuracy. Both the equal weight and performance weighted decision makers show good statistical accuracy with the latter exceeding the informativeness of the former. The results are quite robust against loss of expert and loss of calibration variable. On the whole the expert group is quite coherent. When the sociology of expert judgment comes to be written, this example will be a poster child.

8. References

1. Cooke RM (1991) Experts in Uncertainty-Opinion and Subjective probability in science. (Oxford University Press).

2. Colson AR & Cooke RM (2017) Cross validation for the classical model of structured expert judgment. Reliab. Eng. Syst. Saf. 163:109-120.

3. Colson AR & Cooke RM (2018) Expert Elicitation: Using the Classical Model to Validate Experts' Judgments. Review of Environmental Economics and Policy 12(1):113-132.

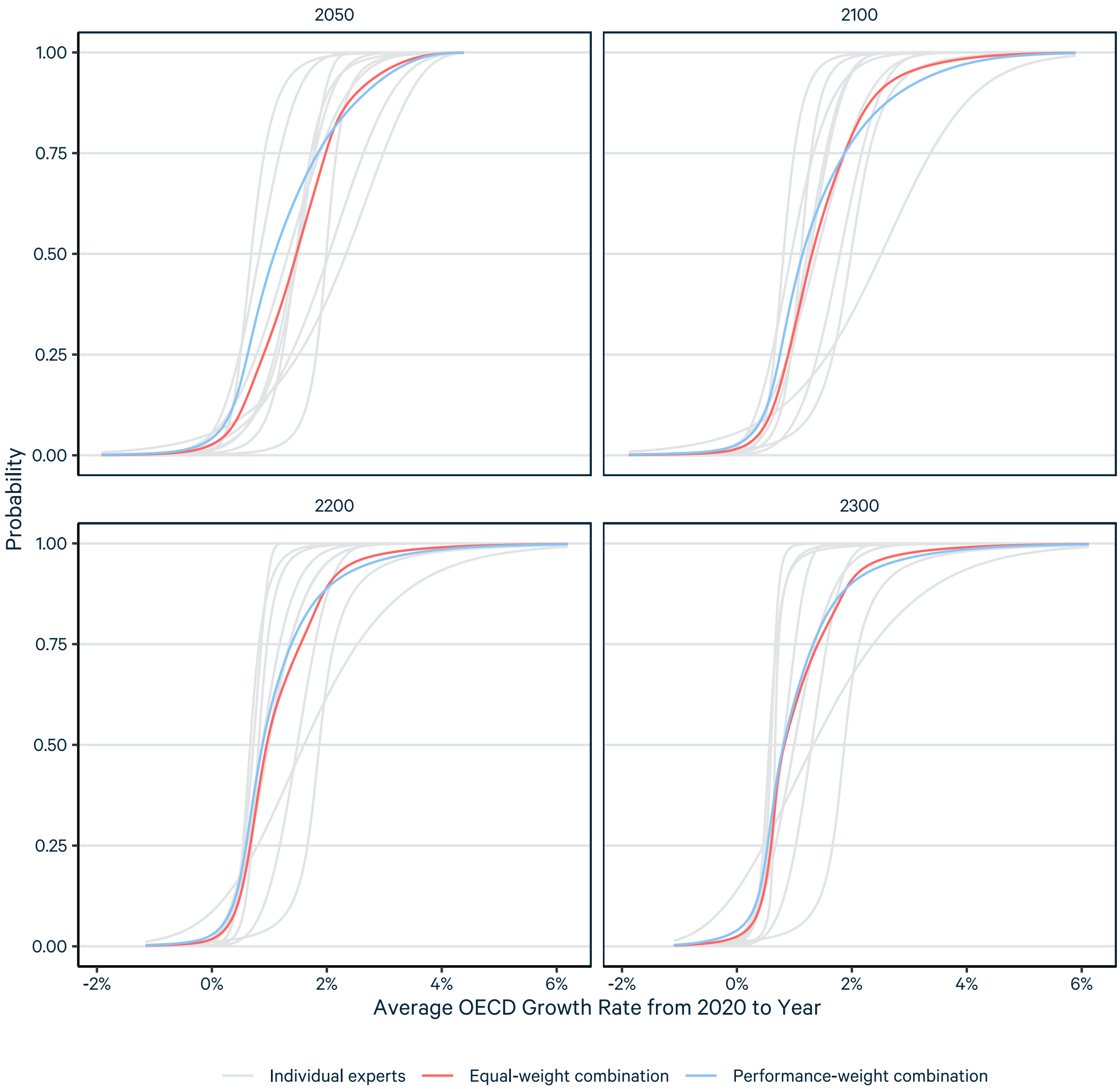
4. Eggstaff JW, Mazzuchi TA, & Sarkani S (2014) The effect of the number of seed variables on the performance of Cooke's classical model. Reliability Engineering & System Safety 121:72-82.

5. Flandoli F, Giorgi E, Aspinall WP, & Neri A (2011) Comparison of a new expert elicitation model with the Classical Model, equal weights and single experts, using a cross-validation technique. Reliability Engineering & System Safety 96(10):1292-1310.

6. Cooke, Roger M., Marti, Deniz and Mazzuchi, Thomas A., (2020) Expert Forecasting with and without Uncertainty Quantification and Weighting: What Do the Data Say? International Journal of Forecasting, published online July 25 2020 **https://doi.org/10.1016/j.ijforecast.2020.06.007**

III. Economic Growth Survey: Additional Results

Figure OA-4. Cumulative Distribution Functions (CDFs) of Average Growth Rates of GDP per Capita for the OECD from the EGS, Individuals and Expert Combinations.



IV. Future Emissions Survey: Additional Description and Results

1. Survey Methodology

To generate very long-run distributions of global emissions of CO2, CH4, and N2O, our Future Emissions Survey (FES) elicited 10 leading experts in socioeconomic projections and climate policy that were nominated by their peers and/or by members of our Scientific Advisory Board. The experts surveyed were based at universities, non-profit research institutions, and multilateral international organizations. They have expertise in, and have undertaken, long-term projections of the energy-economic system under a substantial range of climate change mitigation scenarios.

As with our Economic Growth Survey, the FES employed the Classical Model of structured expert judgment in which experts first quantified their uncertainty with regard to a set of relevant calibration variables for which true values are known, for the purposes of validation and performance weighting in the combining of the expert distributions. Each elicitation was conducted individually by videoconference in July and August of 2021 in sessions that lasted ~2 hours, and experts provided additional detail as needed by email and videoconference. Experts participated in the survey in their own capacity and were provided an honorarium where appropriate.

In the survey, experts provided quantiles of uncertainty (minimum, 5th, 50th, 95th, maximum, as well as additional percentiles at the expert's discretion) for four variables for a case referred to as Evolving Policies, which incorporates views about changes in technology, fuel use, and other conditions, and consistent with the expert's views on the evolution of future policy. The Evolving Policies case corresponds to the USG approach to benefit cost analysis, which evaluates US regulations as incremental against a more expansive backdrop of other policies and conditions and is responsive to NASEM recommendations for including future background policy in the uncertain distributions of socioeconomic projections.

Experts provided quantiles of uncertainty for the following four non-overlapping categories: (1) fossil and process related CO2 emissions; (2) changes in natural CO2 stocks and negative emission technologies; (3) CH4; and (4) N2O. They did this for each benchmark year: 2050, 2100, 2150, 2200, and 2300. For the first category, they were also asked to indicate the sensitivity of emissions to five underlying GDP per capita trajectories. More precisely, they were asked to provide for each benchmark year:

1. Global CO_2 emissions for five levels of future of GDP per capita representing the minimum, 2.5th percentile, 50th percentile, 97.5th percentile, and maximum of projected economic growth in each benchmark year (i.e., a separate set of quantiles for each of the five GDP per capita levels drawn from the MSW

dataset). Reported emissions in this category included net total emissions from processes, including CCS applied to fossil energy and process-related emissions. By construction, emissions reported for this category are greater than or equal to 0.

2. Quantiles of the net CO_2 emissions from the combined sum of Agriculture, Forestry, and Other Land Use (AFOLU) and sequestered emissions from Direct Air Capture (DAC) and Bioenergy with Carbon Capture and Storage (BECCS). By construction, total net emissions in this category could be positive (net CO_2 source) or negative (net CO_2 sink).
3. Quantiles of global CH_4 emissions including CH_4 emissions from AFOLU.
4. Quantiles of global N_2O emissions including N_2O emissions from AFOLU.

The categories were designed, and experts were specifically directed, to avoid double counting of emissions in the distributions provided. Taken together, the elicited source categories account for greater than 95% of current global emissions of greenhouse gases. Experts were permitted, but not required, to provide quantiles for AFOLU, DAC, BECCS, CH_4, and N_2O conditioned on economic growth in the same manner as CO_2 emissions in category 1). Experts were permitted to consult outside sources at their discretion for this section of the survey.

During the survey, experts provided their quantiles by dictating values for their quantiles for each of the specified categories, years, and economic growth trajectories. The values were recorded in a spreadsheet visible to the expert during the elicitation via screen share.

2. Summary of Expert Rationale

As part of FES, experts described their rationale and the conditions supporting their provided distributions of emissions. The rationale were provided independently, but when viewed across the full set of experts, featured a number of common and primary factors, including economic growth, global climate policy ambition and success of implementation, and technology evolution. Each of these factors could work in concert with or against each other to result in the final uncertain distribution.

CO_2 distributions provided for low economic growth scenarios in general incorporated divergent potential outcomes. On one hand, low growth was viewed as providing an impediment to both policy ambition as well as further improvements in and deployment of low- or zero-emission energy technologies. On the other hand, low growth was also generally viewed to reduce global emissions generally as lower economic activity at current or decreased emissions intensity levels would lead to a decrease in emissions overall. A relatively common narrative supporting the higher end of emissions distributions was that low economic growth trajectories could lead to a revisitation of current pledges to reduce emissions and favor continued growth in energy derived from fossil fuels, leading to further lock-in of such technologies.

For median rates of per capita economic growth, experts in general viewed global policy as the primary driver, including the success or failure of countries meeting their pledges under the Paris Agreement and enhancing the ambition of those targets, as well as by the evolution of developing nations' use of fossil fuels and assistance provided. Continued evolution of technology, driven significantly by potential investments consistent with mitigation goals, also featured prominently as a driver. A common result for the 2050 and 2100 the medians of the emissions distributions was a reductions of absolute emissions from today's levels, but with an uncertain range leaving substantial probability for continued, and in some cases quite significant, increases as well. The low emission (5^{th} percentile) quantiles generally represented significant reductions from today's levels but at a level insufficient keep global temperature increases below 1.5 degrees Celsius, even when considering reported quantiles for AFOLU.

For high rates of per capita economic growth, several experts expected that significantly enhanced economic activity would likely lead to increased emissions in the near-term (to 2050 and for some experts to 2100), as the time needed for further development and deployment of zero-emission technologies was insufficient to decrease the emissions intensity (emissions/GDP) quickly enough to offset the economic growth. High economic growth in general was viewed to support increasing attention to reducing emissions from a policy standpoint and an enhancement of global climate policy goals, leading to a more rapid medium- and longer-term transition to greatly reduced emissions overall compared with relatively lower economic growth scenarios. An alternative viewpoint expressed was that greater wealth could also allow for greater adaptation or indifference to the effects of climate change, thereby acting as a brake on policy ambition and allowing for continued increases in emissions well into the future.

Several experts also observed that, if global policy were to remain largely centered around absolute quantity targets (e.g., percent reduction from 2005 levels or net zero by a date certain), emissions would be relatively decoupled from economic growth. Their view of this decoupling manifested itself in the form of relatively low variation between their distributions across economic growth trajectories. Similarly, some of the experts felt that the high economic growth trajectories in particular represented worlds in which economic growth was decoupled from emissions.

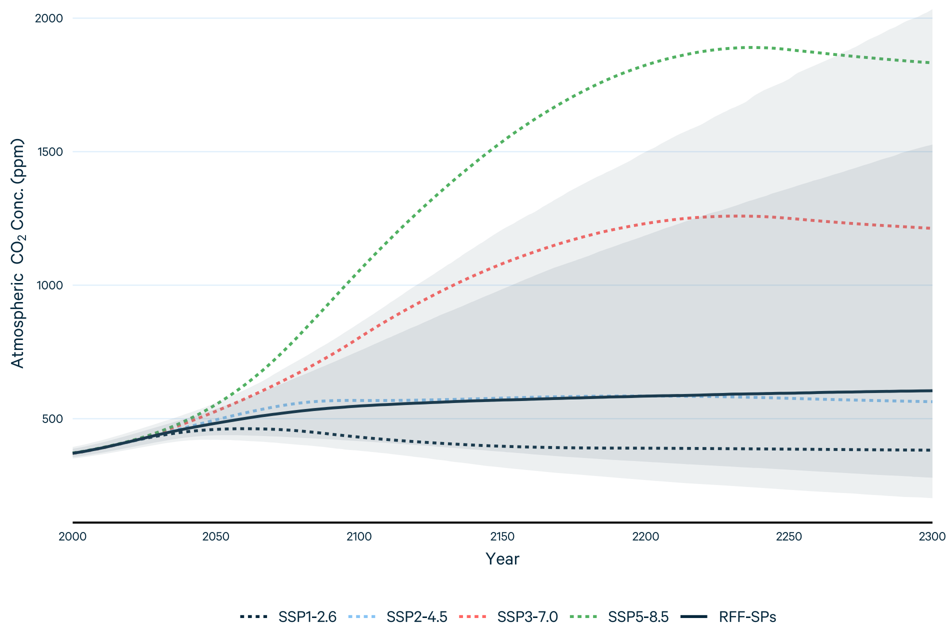
Experts generally viewed near-term potential from DAC and BECCS as limited through 2050, but they became an increasing and more substantial part of the solution alongside natural land sinks by 2100. In general the experts allowed room for the narrative that society may wish to have net negative annual emissions for several decades even after eliminating direct emissions to draw down atmospheric CO2 concentrations to return to a level consistent with current or previous levels.

Emissions from methane are primarily driven by livestock (enteric fermentation), agriculture (including cultivation of rice), and fossil fuels (natural gas), and experts' distributions were primarily driven by their expectations on the future evolution of emissions from these sources. Experts that viewed a rapid transition to zero-emitting energy sources as relatively likely tended to have distributions that reflected the rapid zeroing out of the component of CH4 emissions from fossil fuels. Other experts allowed for the expansion of such emissions in acknowledgement that natural gas may be relied upon heavily as a transition fuel or even as a significant and substantially increasing source of uncontrolled emissions over the long-term. In nearly all cases there was a non-zero lower limit reached, even in the lower quantiles, in acknowledgment that some components of these emissions were unlikely to be fully eliminated by complete modification of diet or agricultural practices. Experts' supporting rationale for N2O, which is similarly associated with agriculture and dietary preferences, generally followed a narrative consistent with the corresponding elements from CH_4.

During the FES, some experts expressed a desire for further control over the correlations between variables. The first correlation desired was to condition emissions on population in addition to GDP per capita. Experts expressing this desire in general agreed with the design decision to condition on GDP per capita rather than GDP (the product of GDP per capita and population) as the primary variable and accommodated the study design by providing quantiles constructed to incorporate the possibility of low and high population futures. Some experts also expressed a desire to more tightly couple their quantiles of category 1 emissions with the potentially negative emissions from category 2 by providing a single distribution of net emissions from both categories. These experts generally viewed the level of negative emissions as being tailored to achieve a particular atmospheric CO_2 outcome or directly to offset emissions. Accounting for this correlation would lead to narrower distributions of net emissions overall.

3. Additional Results for CO_2 Emissions

Figure OA-5. Projected total atmospheric CO_2 concentration.



Notes: Solid lines represent median values, dark and light shading represent the 5th to 95th (darker) and 1st to 99th (lighter) percentile ranges based on the RFF-SPs.

Figure OA-6. Cumulative distribution functions (CDFs) of individual and combined expert projections for annual CO_2 emissions across a range of timeframes and GDP per capita growth trajectories

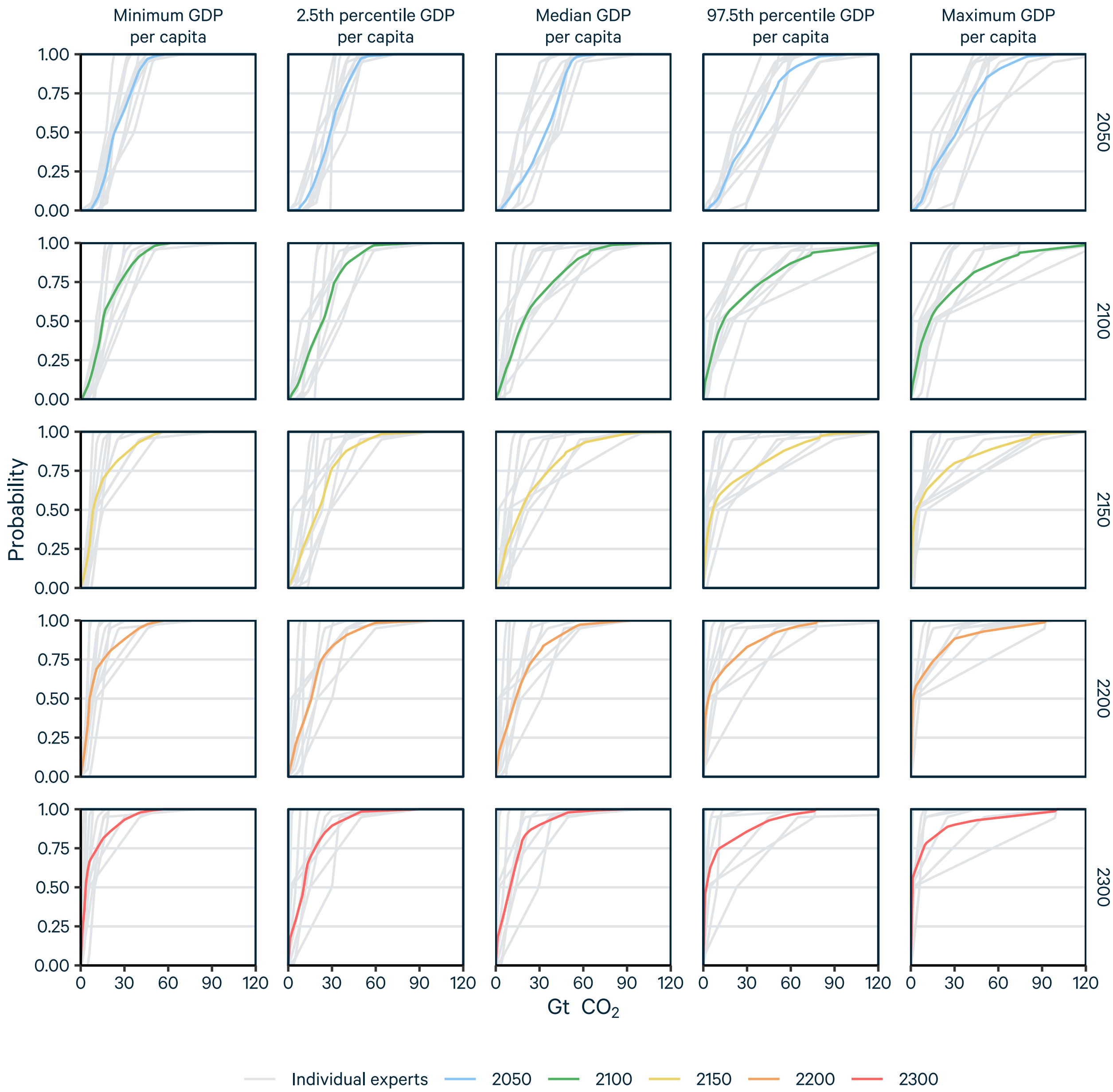


Figure OA-7. Annual net CO_2 emissions from natural carbon stocks and negative emissions technologies

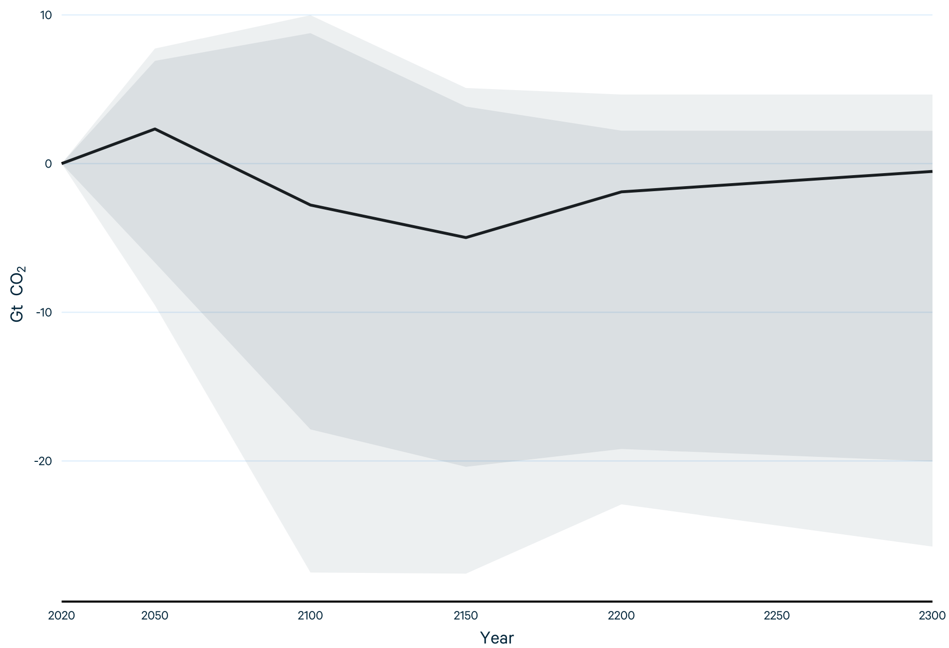
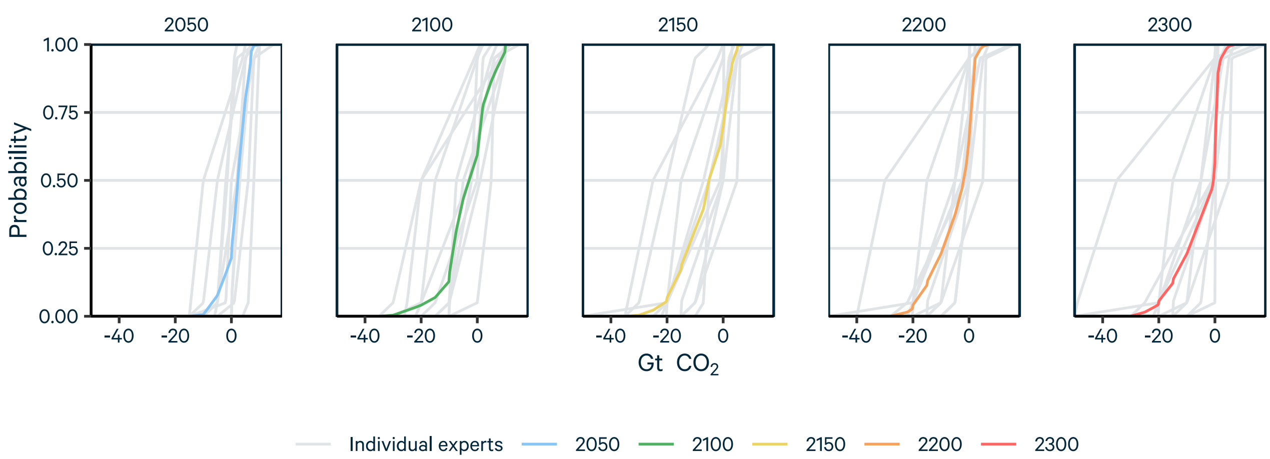
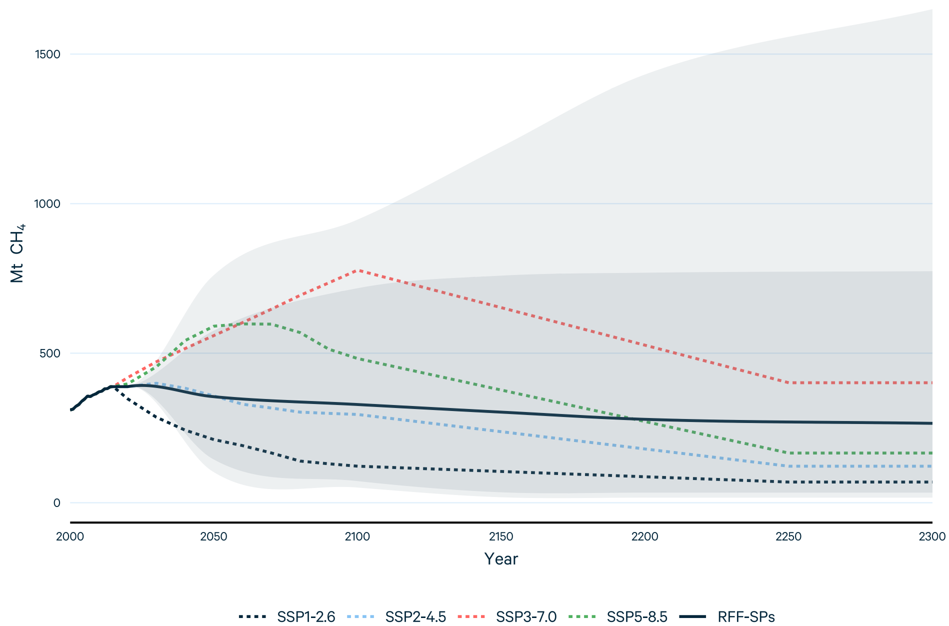


Figure OA-8. Cumulative distribution functions (CDFs) of individual and combined expert projections for net CO_2 emissions from natural carbon stocks and negative emissions technologies across a range of timeframes



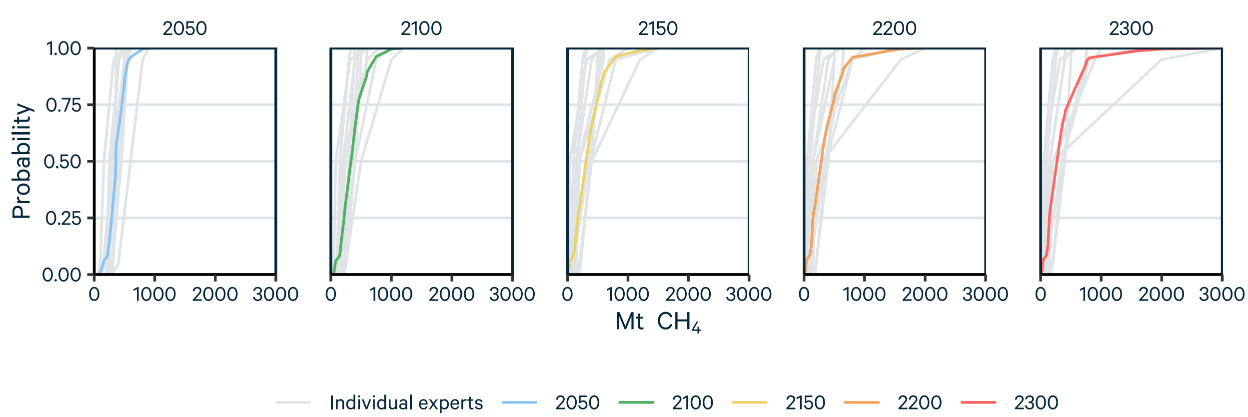
5. CH_4 Emissions

Figure OA-9. Annual emissions of CH_4 from the RFF-SPs and the SSPs.

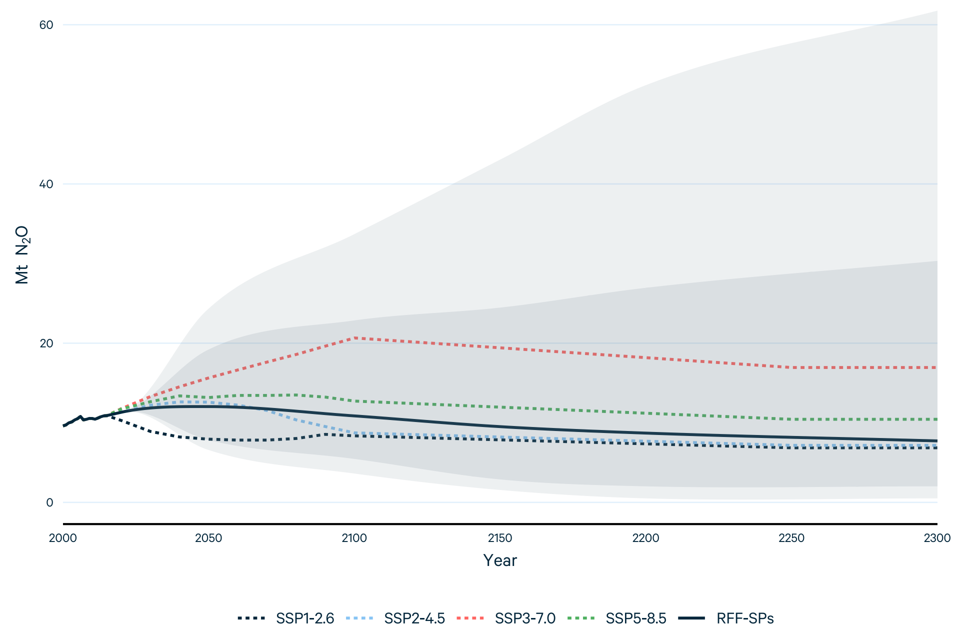


Lines represent median values, and dark and light shading represent the 5th to 95th (darker) and 1st to 99th (lighter) percentile ranges of the RFF-SPs.

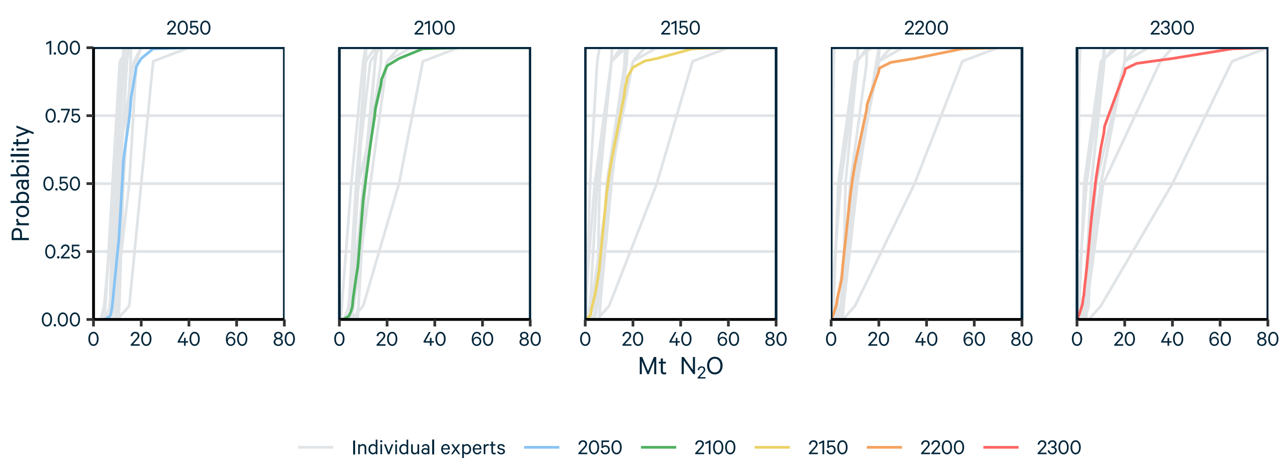
Figure OA-10.Cumulative distribution functions (CDFs) of individual and combined expert projections for annual CH_4 emissions across a range of timeframes



6. N_2O emissions

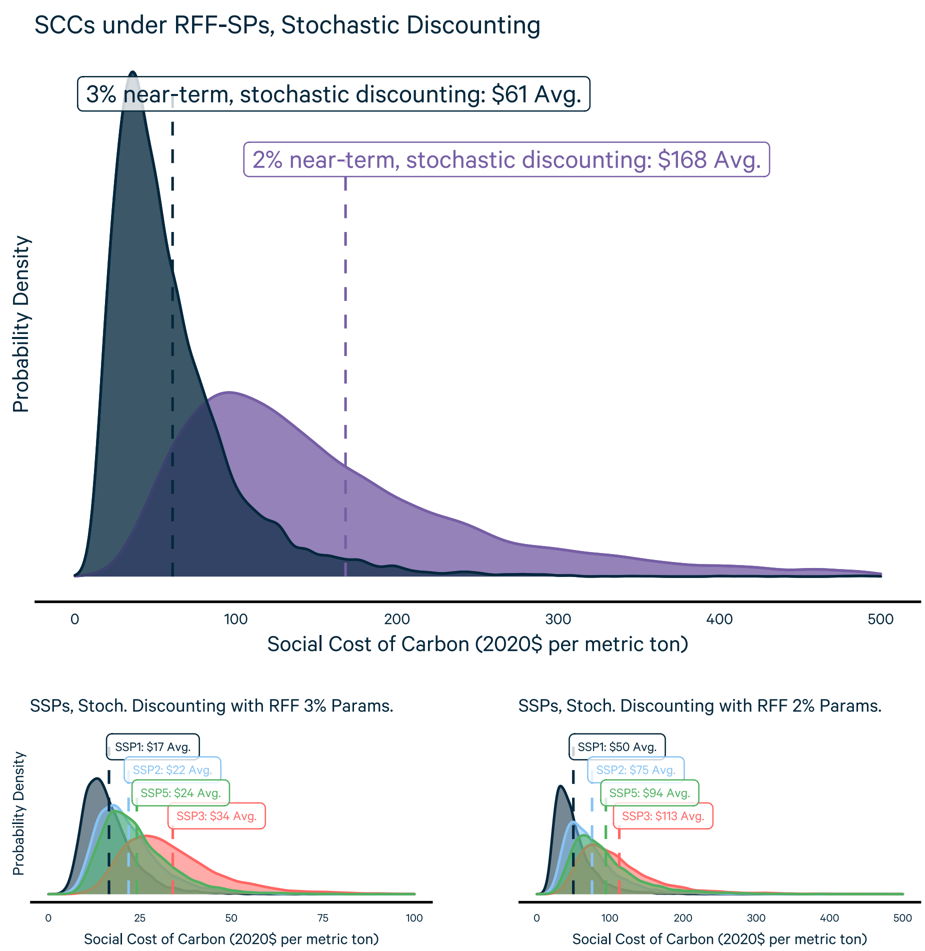


Notes: Lines represent median values, and dark and light shading represent the 5th to 95th (darker) and 1st to 99th (lighter) percentile ranges of the RFF-SPs.



V. Additional SCC Calculations

Figure OA-13. Illustrative Probability Distributions of the Social Cost of Carbon (2020$/ton CO_2) with FaIR Climate and DICE Damage Modules, under Alternative Socioeconomic Inputs, using our Stochastic Discounting Parameters for All Socioeconomics ($\rho = 0.8\%, \eta = 1.57$ for 3% near-term, or $\rho = 0.2\%, \eta = 1.24$ for 2% near-term)



VI. Economic Growth Survey Elicitation Protocol

Economic Growth Survey

Overview:

Resources for the Future (RFF)is conducting its *Economic Growth Survey* to implement National Academy of Sciences (NAS) **recommendations** to improve the long-run economic growth projections that support estimates of the Social Cost of Carbon (SCC). The SCC is an economic metric used by the US federal government, state governments, and foreign governments to account for climate change in their actions. RFF's *Economic Growth Survey* is being carried out as a part of RFF's Social Cost of Carbon **initiative.**

- To implement the survey, RFF is conducting individual elicitations of ~15 leading experts in economic growth that have been nominated by their peers.
- Each expert will provide quantiles of future mean GDP per capita levels for the OECD as well as for a number of other regions, for the periods 2015-2050, 2015-2100, 2015-2200, and 2015-2300.
- Each expert will additionally quantify his or her uncertainty with regard to a set of relevant calibration variables for which true values are known.
- RFF will report combined distributions of economic growth projections based upon equal-weight combinations of the distributions provided by the experts as well as performance-weighted combinations, with performance measured via the calibration questions.
- By design, each expert achieves his or her maximal long run expected combined score on the calibration questions by, and only by, stating percentiles corresponding to his or her true beliefs.
- Expert names are preserved to enable competent peer review but are not associated with responses in any published documentation. Expert reasoning is captured during the elicitation and becomes, where indicated, part of the published record.
- Each elicitation will take approximately 2 hours and be conducted over freely available videoconferencing software. Experts receive an honorarium of $1250.

The elicitation comprises three parts:

Part 1: The expert will answer four practice questions for which, after responding, the answers will be provided to the expert. These questions are intended to orient the experts to the elicitation methodology as well as to increase their awareness to potential biases and overconfidence.

Part 2: The expert will answer 11 calibration questions that are intended to be used to performance-weight the experts when combining the full set of elicitation results on the main variables of interest.

Part 3: The expert will report values for the variables of interest for this study: quantiles of mean OECD GDP per capita levels for OECD countries, as well as for a number of other regions, for the periods 2015-2050, 2015-2100, 2015-2200, and 2015-2300.

Additional Detail

Resources for the Future's Social Cost of Carbon (SCC) **initiative** is implementing National Academy of Sciences (NAS) **recommendations** to improve the socioeconomic projections (population, economic growth, global emissions) that underpin estimates of the SCC, the economic metric used by the US federal government, state governments, and foreign governments to account for climate change in their actions.

The NAS recommended that, in order to provide for the long residence time of carbon dioxide in the atmosphere, the SCC model runs (and the socioeconomic projections required) should extend far enough into the future to account for the vast majority of resulting damages (i.e. multiple centuries). The NAS further recommended that the appropriate way to address challenges inherent in quantifying the uncertainty for projections for such an extended time horizon is to use statistical methods in concert with results based upon the formal elicitation of experts, referred to as structured expert judgment. RFF's program of research, under which this structured expert judgment on economic growth falls, is intended to implement these recommendations and yield a set of very long-run (multi-century) central projections for economic growth, population, and global emissions, with associated uncertainty bounds.

Structured Expert Judgment Methodology

The goal of structured expert judgment is to generate a probability distribution for one or more variables of interest by combining a set of individual distributions for the variables that have been provided by a set of experts. There are a number of ways to generate the resulting combined distribution, the simplest being to combine the set of individual distributions in equal weights. An alternate method, which generally provides advantages of narrower overall uncertainty distributions with greater statistical accuracy and has been shown to provide greater performance both **in-sample** and **out-of-sample**, performance-weights the experts according to their ability to quantify their uncertainty for a set of calibration variables for which the true values are known. This latter approach is exemplified by the **Classical Model** for structured expert judgment, so called for its analogy with classical hypothesis testing. In its publication of its research on this topic, RFF will discuss distributions generated based upon both approaches.

In the Classical Model, each of the experts on a panel quantifies his or her uncertainty with regard to variables of interest as well as with regard to a set of calibration variables from the subject area for which true values are known. Experts are treated as statistical hypotheses and scored on two performance metrics -- *statistical accuracy* and *informativeness* -- on the calibration variables.

Statistical accuracy: Roughly, an expert is statistically accurate if, in a statistical sense, 5% of the true values fall below his/her 5th percentiles, 45% of the realizations fall between his/her 50th and 5th percentiles, etc. More formally, the statistical accuracy of a given expert is measured as the probability (P-Value) of falsely rejecting the hypothesis that an expert's observed inter-percentile frequencies comply with his/her probabilistic assessments.

Informativeness: The informativeness of an expert for a given question is related to the width of the uncertain distribution provided by the expert for that question. Narrower error bounds will yield a higher score for informativeness relative to wider error bounds. More formally, the informativeness of an expert is measured as Shannon relative information with respect to a background measure. Per variable, the background measures are uniform on an interval containing all assessments, with an analyst-stipulated small overshoot.

An expert's overall performance for a given question is based upon the product of his or her statistical accuracy and informativeness. Taking the product of these two metrics has the important property that it results in an *asymptotically strictly proper scoring rule*. In practice, this means that an expert achieves his or her maximal long run expected combined score by and only by stating percentiles corresponding to his or her true beliefs. Statistical accuracy is a fast function, typically varying over several orders of magnitude in a typical panel of experts. Informativeness is a slow function typically varying within a factor 3. Normalizing the combined scores of weighted experts allows statistical accuracy to dominate with informativeness modulating between experts of comparable accuracy.

Part 1: Practice questions

- *For each of the following values requested, please provide your 5th, 50th, and 95th percentiles. In other words: for the 5th percentile, provide the value for which the true value has a 1 in 20 chance of being less than; for the 50th percentile provide the value for which the true value has an equal chance of being greater or less than; and for the 95th percentile provide the value for which the true value has a 1 in 20 chance of exceeding.*
- *Please refrain from consulting outside sources, including those found on the internet, in answering these questions.*

The geometric mean of annual growth rates yields the rate at which constant growth over the full period would result in the observed changes in GDP levels from the beginning of the period to the end – e.g.: $GDP_{2017} = GDP_{1980} * ((1 + \text{geometric mean})^{38})$.

The geometric mean of the **global** annual growth rate of GDP from 1980 to 2017 was 2.86% per year.

1. **What is the geometric mean of the annual growth rate from 1980 to 2017 for Saudi Arabia?**

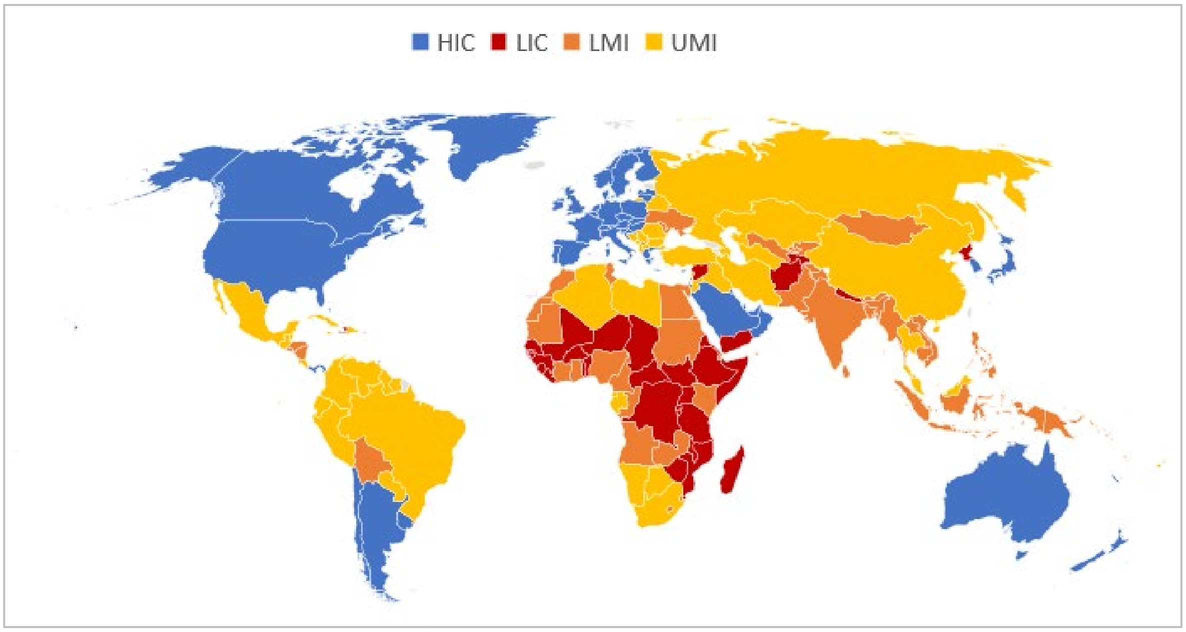
Also for Saudi Arabia, we are interested in your assessment of the **variability** of annual economic growth. One way of measuring such variability is by evaluating the Mean Absolute Deviation (MAD) with respect to the region's arithmetic mean for the full period.

$$MAD = \frac{1}{38} \sum_{y=1980}^{2017} |region\ growth\ rate\ in\ year\ y - \ mean\ region\ growth\ rate\ |$$

2. **What is the MAD of annual GDP growth for Saudi Arabia from 1980 to 2017?**

The following questions assess a metric of competitiveness and economic activity for two country groupings: Low Income Countries (LIC) and High Income Countries (HIC).

Note that LIC and HIC are two of four total income categories, along with Lower Middle Income (LMI) and Upper Middle Income (UMI). Appendix A provides a full list of countries in each category.



The *time required to start a business* -- the number of calendar days needed to complete the procedures to legally operate a business -- is a metric of competitiveness surveyed annually on a global basis.

- The survey conducted:
    - reflects the time for a small- to medium-size limited liability company to start up and formally operate in each economy's largest business city;
    - uses a standardized business that is 100% domestically owned, has a start-up capital equivalent to 10 times the income per capita, engages in general industrial or commercial activities and employs between 10 and 50 people one month after the commencement of operations, all of whom are domestic nationals;
    - considers two cases of local limited liability companies that are identical in all aspects, except that one company is owned by five married women and the other by five married men.
- If a procedure can be speeded up at additional cost, the fastest procedure, independent of cost, is chosen.

The full, detailed methodology for the survey will be provided at the request of the expert.

The unweighted country average for the world for number of days to start a business went from **51.55 days** in **2003** to **20.12 days** in **2018**.

3. What is the average number of days to start a business in 2018 for LIC?

4. What is the average number of days to start a business in 2018 for HIC?

Part 2: Performance calibration questions

- *For each of the following values requested, please provide your 5th, 50th, and 95th percentiles. In other words: for the 5th percentile, provide the value for which the true value has a 1 in 20 chance of being less than; for the 50th percentile provide the value for which the true value has an equal chance of being greater or less than; and for the 95th percentile provide the value for which the true value has a 1 in 20 chance of exceeding.*
- *Please refrain from consulting outside sources, including those found on the internet, in answering these questions.*

The geometric mean of annual growth rates yields the rate at which constant growth over the full period would result in the observed changes in GDP levels from the beginning of the period to the end. – e.g.: $GDP_{2017} = GDP_{1980} * ((1 + \text{geometric mean})^{38})$.

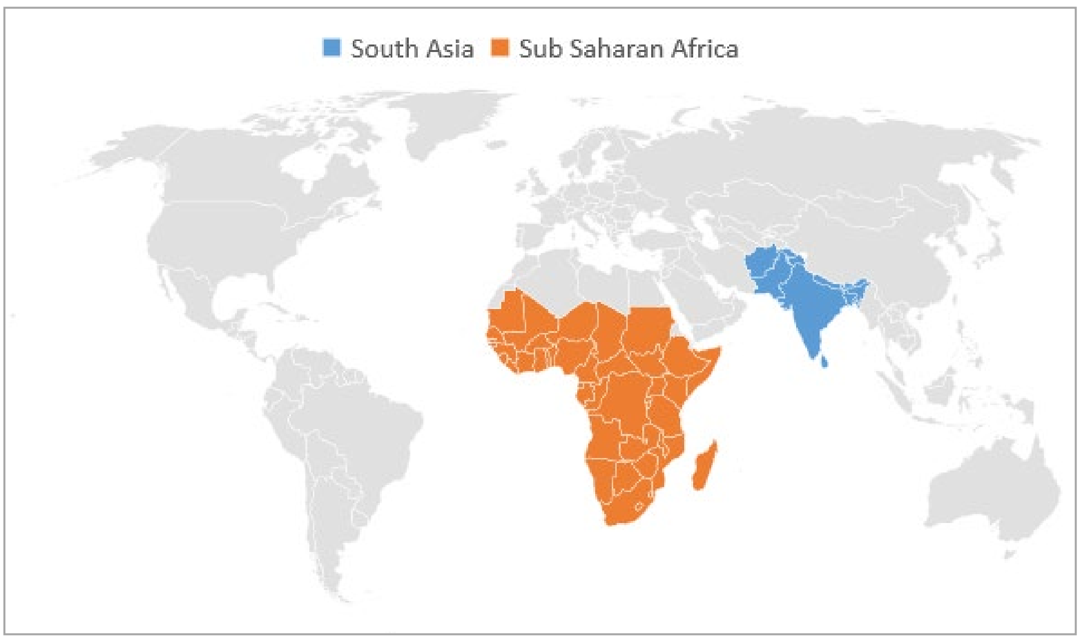
The geometric mean of the **global** annual growth rate of GDP from 1980 to 2017 was 2.86% per year.

1. **What is the geometric mean of the annual growth rate from 1980 to 2017 for China?**

2. **What is the geometric mean of the annual growth rate from 1980 to 2017 for South Asia*?**

3. **What is the geometric mean of the annual growth rate from 1980 to 2017 for Sub-Saharan Africa*?**

*Full list of countries in South Asia and Sub-Saharan Africa is included in Appendix B.



For the same set of regions, we are interested in your assessment of the *variability* of annual economic growth.

One way of measuring such variability is by evaluating the Mean Absolute Deviation (MAD) with respect to a country or region's arithmetic mean for the full period.

$$MAD = \frac{1}{38}\sum_{y=1980}^{2017} |region\ growth\ rate\ in\ year\ y - \ mean\ region\ growth\ rate\ |$$

4. What is the MAD of annual GDP growth for China from 1980 to 2017?

5. What is the MAD of annual GDP growth for South Asia from 1980 to 2017?

6. What is the MAD of annual GDP growth for Sub-Saharan Africa from 1980 to 2017?

For four decades, the US Congressional Budget Office (CBO) has prepared economic forecasts for use in making its projections for the federal budget.

- The CBO prepares *next-year* forecasts in January of a given year which forecast the geometric mean of growth from the beginning of the given year through the end of the following year.
- The CBO also reports *5-year* forecasts in January of a given year which forecast the geometric mean of growth from the beginning of the given year through the end of the following four years.

To evaluate its economic forecasts, CBO compares them with the economy's actual performance. The error of each forecast is measured as an **absolute** difference between the forecasted growth and the actual growth. The following question requests comparison of the accuracy of CBO's next-year forecasts with the accuracy of CBO's 5-year forecasts.

The average *absolute* error of **next-year** forecasts of GDP growth rate in the United States from 1992 to 2014, calculated according to the formula below, was **1.08%.**

$$\frac{1}{23}\sum_{t=1992}^{2014}\left|(avg^{*}\ growth\ forecast\ from\ t\ to\ t+1)-\ (avg^{*}\ growth\ from\ t\ to\ t+1)\right|$$

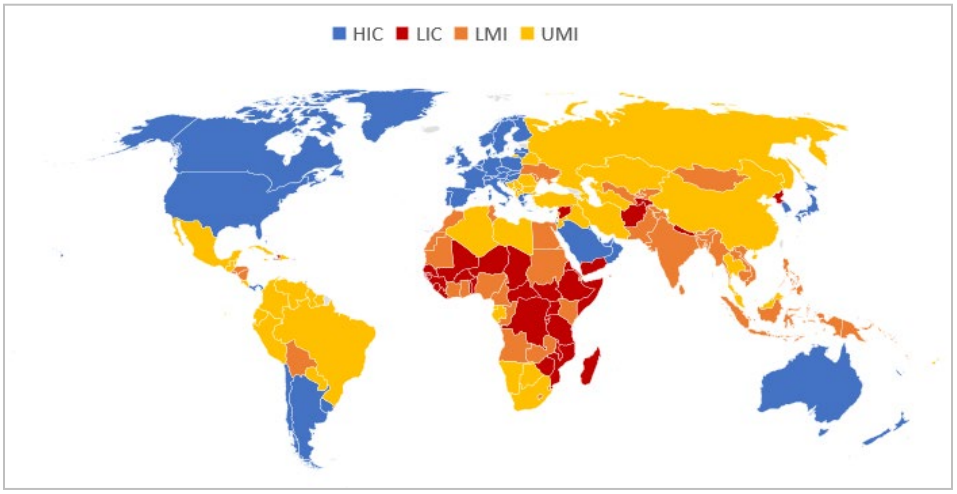
7. What is the average absolute error of CBO's 5-year forecasts of GDP growth rate in the United States from 1992 to 2011?

$$\frac{1}{20}\sum_{t=1992}^{2011}\left|(avg^{*}\ growth\ forecast\ from\ t\ to\ t+4)-\ (avg^{*}\ growth\ from\ t\ to\ t+4)\right|$$

*****averages calculated as a geometric mean of growth

As in the practice questions, the following questions request assessment of metrics of competitiveness and economic activity for two groupings of countries: Low Income Countries (LIC) and High Income Countries (HIC).

Note that LIC and HIC are two of four total income categories, along with Lower Middle Income (LMI) and Upper Middle Income (UMI). Appendix A provides full list of countries in each category.



The *average cost of starting a business*, recorded as a percentage of the economy's Gross National Income per capita (GNIpc), is surveyed annually on a global basis.

In the survey:

- The representative business is a small- to medium-size limited liability company (with details as previously described in Part I) commencing formal operations in each economy's largest business city;
- Costs includes all official fees and fees for legal or professional services if such services are required by law or commonly used in practice;
- Fees for purchasing and legalizing company books are included if these transactions are required by law;
- Although value added tax registration can be counted as a separate procedure, value added tax is not part of the incorporation cost;
- The company law, the commercial code and specific regulations and fee schedules are used as sources for calculating costs; and
- In all cases the cost excludes bribes.

The **unweighted world average** cost of starting a business as a percentage of GNIpc went from **104.68%** in **2003** to **23.88%** in **2018**.

8. **What was the average cost of starting a business as percentage of GNIpc in 2018 for LIC?**

9. **What was the average cost of starting a business as percentage of GNIpc in 2018 for HIC?**

The *time required to enforce a contract* -- the number of calendar days from the filing of a lawsuit in court until the final determination and, in appropriate cases, payment -- is surveyed annually on a global basis.

In the representative case employed by the survey:

- The case is a commercial dispute between a buyer and a seller that is resolved through a local first-instance court;
- The value of the claim is equal to 200% of the economy's income per capita or $5,000, whichever is greater.
- The dispute involves custom-tailored goods that are rejected by the buyer as being of insufficient quality.
- The seller disputes the claim in court and the judge renders a judgment that is 100% in favor of the seller.
- The buyer does not appeal the judgment. The seller decides to start enforcing the judgment as soon as the time allocated by law for appeal lapses.
- The seller takes all required steps for prompt enforcement of the judgment. The money is successfully collected through a public sale of the buyer's movable assets (for example, office equipment and vehicles).

The **unweighted world average** for time to enforce a contract went from **605.41 days** in **2003** to **647.54 days** in **2018**.

10. **What was the average number of days to enforce a contract in 2018 for LIC?**

11. **What was the average number of days to enforce a contract in 2018 for HIC?**

Part 3: Elicitation

We will now elicit your 1st, 5th, 50th, 95th, and 99th percentiles on *GDP per capita levels for OECD countries* in 2050, 2100, 2200, and 2300.

- As in the calibration questions, your 50th percentile is the value for which you believe the true value has an equal chance of being less or greater than. Your 1st and 99th percentiles are the values for which you believe the true value has a 1 in 100 chance of being less or greater than, respectively. Your 5th and 95th percentiles are the values for which you believe the true value has a 1 in 20 chance of being less or greater than, respectively.
- Many economists think primarily in terms of growth *rates* rather than levels, so we are providing a spreadsheet tool for translating between an average growth rate over a period and the resulting GDP per capita level at the end of the period.
- Given that the final variables of interest are the levels of GDP per capita, we encourage you to consciously pay attention to the levels that result from changing the growth rates in the table.
- As a part of the elicitation we will also ask you to describe your rationale for the quantiles. For example, we will ask you to describe future narratives that would plausibly yield the reported levels of GDP per capita with the likelihoods indicated.
- We will also ask you to identify the primary drivers of your low and high quantiles by selecting them from a list of relevant factors that have been suggested to us by experts as well as drawn from the growth literature. Experts are encouraged to add their own primary factors to the list if they are not included in the initial list of drivers.
- In order to capture your rationale, we will be taking notes throughout the elicitation. To facilitate such notetaking and ensure its veracity, we will also request that this section be recorded, solely for our internal use in processing the results. *Such recording is completely optional and at the discretion of the individual expert.*
- *Consulting outside sources for this part of the elicitation is permitted.*

Guide to the elicitation reporting spreadsheet:

1) The first tab in the spreadsheet, labeled ‘Rationale’, provides the list of potential factors that have been suggested by experts consulted and the growth literature to influence long-run economic growth.
2) The second tab, labeled ‘OECD’, provides the tool with which the expert can enter quantiles for growth rates to convert them into levels for final reporting. This tab additionally provides different views of the historic OECD GDP per capita levels and growth rates from 1915 to 2014.

Appendix A - Income Categories

HIC (80 countries)	LIC (34 countries)	LMI (47 countries)	UMI (56 countries)
Aruba	Afghanistan	Angola	Albania
Andorra	Burundi	Bangladesh	Armenia
United Arab Emirates	Benin	Bolivia	American Samoa
Argentina	Burkina Faso	Bhutan	Azerbaijan
Antigua and Barbuda	Central African Republic	Cote d'Ivoire	Bulgaria
Australia	Congo, Dem. Rep.	Cameroon	Bosnia and Herzegovina
Austria	Comoros	Congo, Rep.	Belarus
Belgium	Eritrea	Cabo Verde	Belize
Bahrain	Ethiopia	Djibouti	Brazil
Bahamas, The	Guinea	Egypt, Arab Rep.	Botswana
Bermuda	Gambia, The	Micronesia, Fed. Sts.	China
Barbados	Guinea-Bissau	Georgia	Colombia
Brunei Darussalam	Haiti	Ghana	Costa Rica
Canada	Liberia	Honduras	Cuba
Switzerland	Madagascar	Indonesia	Dominica
Channel Islands	Mali	India	Dominican Republic
Chile	Mozambique	Kenya	Algeria
Curacao	Malawi	Kyrgyz Republic	Ecuador
Cayman Islands	Niger	Cambodia	Fiji
Cyprus	Nepal	Kiribati	Gabon
Czech Republic	Korea, Dem. People's Rep.	Lao PDR	Equatorial Guinea
Germany	Rwanda	Sri Lanka	Grenada
Denmark	Senegal	Lesotho	Guatemala
Spain	Sierra Leone	Morocco	Guyana
Estonia	Somalia	Moldova	Iran, Islamic Rep.
Finland	South Sudan	Myanmar	Iraq
France	Syrian Arab Republic	Mongolia	Jamaica
Faroe Islands	Chad	Mauritania	Jordan
United Kingdom	Togo	Nigeria	Kazakhstan
Gibraltar	Tajikistan	Nicaragua	Lebanon
Greece	Tanzania	Pakistan	Libya
Greenland	Uganda	Philippines	St. Lucia
Guam	Yemen, Rep.	Papua New Guinea	Maldives
Hong Kong SAR, China	Zimbabwe	West Bank and Gaza	Mexico
Croatia		Sudan	Marshall Islands
Hungary		Solomon Islands	Macedonia, FYR
Isle of Man		El Salvador	Montenegro
Ireland		Sao Tome and Principe	Mauritius
Iceland		Swaziland	Malaysia
Israel		Timor-Leste	Namibia

Italy
Japan
St. Kitts and Nevis
Korea, Rep.
Kuwait
Liechtenstein
Lithuania
Luxembourg
Latvia
Macao SAR, China
St. Martin (French part)
Monaco
Malta
Northern Mariana Islands
New Caledonia
Netherlands
Norway
New Zealand
Oman
Panama
Palau
Poland
Puerto Rico
Portugal
French Polynesia
Qatar
Saudi Arabia
Singapore
San Marino
Slovak Republic
Slovenia
Sweden
Sint Maarten (Dutch part)
Seychelles
Turks and Caicos Islands
Trinidad and Tobago
Uruguay
United States
British Virgin Islands
Virgin Islands (U.S.)

Tunisia
Ukraine
Uzbekistan
Vietnam
Vanuatu
Kosovo
Zambia

Nauru
Peru
Paraguay
Romania
Russian Federation
Serbia
Suriname
Thailand
Turkmenistan
Tonga
Turkey
Tuvalu
St. Vincent and the Grenadines
Venezuela, RB
Samoa
South Africa

Appendix B

South Asia

Afghanistan
Bangladesh
Bhutan
India
Sri Lanka
Maldives
Nepal
Pakistan

Sub-Saharan Africa

Angola
Burundi
Benin
Burkina Faso
Botswana
Central African Republic
Cote d'Ivoire
Cameroon
Democratic Republic of the Congo
Congo
Comoros
Cabo Verde
Eritrea
Ethiopia
Gabon
Ghana
Guinea
Gambia
Guinea-Bissau
Equatorial Guinea
Kenya
Liberia
Lesotho
Madagascar
Mali
Mozambique
Mauritania
Mauritius
Malawi
Namibia
Niger
Nigeria
Rwanda
Sudan
Senegal
Sierra Leone
Somalia
South Sudan
Sao Tome and Principe
Swaziland
Seychelles
Chad
Togo
United Republic of Tanzania
Uganda
South Africa
Zambia
Zimbabwe

Request for additional information

In the projections you previously provided as a part of RFF's Economic Growth Survey, both the economic effects of **physical damages** resulting from future climate change as well as the effects of **policies to address climate change** on economic growth may have played a role in your assessment of future growth paths. Two additional expected use cases for the final, aggregate growth projections would benefit from quantifying the contribution of such effects to your projections to the extent possible.

Please respond to the following pair of questions:

1. Would specifically **excluding** the **economic effects of physical damages** resulting from future climate change significantly alter the economic growth quantiles you have already provided? If so, please use the spreadsheet provided to modify your previous quantiles to exclude such effects.

- The specific thought exercise by which to consider this question is to imagine that actions continue to be taken to address climate change throughout the period, but that in retrospect:
    - Temperature, precipitation, and other physical climate variables did not change in response to increased CO2 levels;
    - Increased levels of atmospheric carbon dioxide did not affect the economy through other physical pathways (e.g. ocean acidification, CO2 fertilization, etc).

2. Would specifically **excluding the economic effects of physical damages** resulting from future climate change (as in question 1) **AND** the **economic effects of future policies** to address climate change significantly alter the economic growth quantiles you have already provided? If so, could you please use the spreadsheet provided to modify your previous quantiles to exclude such effects.

- The specific thought exercise by which to consider this question is to imagine that:
    - at the start of the growth period, it is proven definitively that (as in 1) increases in greenhouse gas emissions will not result in changes in temperature, precipitation, or other physical climate variables. Nor will it lead to effects on other environmental variables (e.g. ocean acidification, plant growth, etc).
    - As a result, throughout the period, no policy actions are taken on the basis of expected climate change.

VII. Future Emissions Survey Elicitation Protocol

RFF Future Emissions Survey

Overview:

Resources for the Future (RFF) is conducting its *Future Emissions Survey* to implement National Academy of Sciences (NAS) **recommendations** to improve the long-run economic growth projections that support estimates of the Social Cost of Carbon (SCC). The SCC is an economic metric used by the US federal government, state governments, and foreign governments to account for climate change in their actions. RFF's *Future Emissions Survey* is being carried out as a part of RFF's Social Cost of Carbon **initiative.**

- To implement the survey, RFF is conducting individual elicitations of ~12 leading experts in socioeconomic projections and climate policy that have been nominated by their peers.
- Each expert will provide quantiles of future emissions for carbon dioxide (CO_2), Methane (CH_4), and Nitrous Oxide (N_2O) and related variables for the years 2050, 2100, 2150, 2200, and 2300.
- Each expert will additionally quantify his or her uncertainty with regard to a set of relevant calibration variables for which true values are known.
- RFF will report combined distributions of emissions projections based upon equal-weight combinations of the distributions provided by the experts as well as performance-weighted combinations, with performance measured via the calibration questions.
- By design, each expert achieves his or her maximal long run expected combined score on the calibration questions by, and only by, stating percentiles corresponding to his or her true beliefs.
- Expert names and qualifications are part of the public record. The association of names and information provided is preserved to enable competent peer review but is not part of any published documentation. Expert reasoning is captured during the elicitation and becomes, where indicated, part of the published record.
- Each elicitation will take approximately 2 hours and be conducted over freely available videoconferencing software. Experts receive an honorarium of $1000.

The elicitation comprises three parts:

Part 1: The expert will answer two practice questions for which, after responding, the answers will be provided to the expert. These questions are intended to orient the experts to the elicitation methodology as well as to increase their awareness to potential biases and overconfidence.

Part 2: The expert will answer 11 calibration questions that are intended to be used to performance-weight the experts when combining the full set of elicitation results on the main variables of interest.

Part 3: The expert will report values for the variables of interest for this study as described on the next page.

Description of information provided by the experts

Experts will be asked to provide quantiles of uncertainty (minimum, 5th, 50th, 95th, maximum, as well as additional percentiles at the expert's discretion) for several variables for the following two cases:

- **Evolving Policies**: Incorporating expected changes in technology, fuel use, and other conditions, *consistent with the expert's expected evolution of future policy*.
- **Current Laws and Regulations**: Incorporating expected changes in technology, fuel use, and other conditions, *consistent with current on-the-books policies*.
    - Emissions distributions offered in under this case should represent current legislation and environmental regulations, including recent government actions for which implementing regulations were available as of August 2, 2021. The potential effects of proposed legislation, regulations, and standards—or sections of legislation that have been enacted but require funds to execute or do not have the required implementing regulations in place—should not be reflected here.

For each case, experts will be asked to provide quantiles of uncertainty for the following for the years 2050, 2100, 2150, 2200, 2300:

1) Global CO_2 emissions ($GtCO_2$) for 3 future trajectories of GDP per capita representing the 2.5th percentile, 50th percentile, and 97.5th percentile of projected economic growth (three separate responses). Experts will also specify distributions for the minimum and maximum GDP per capita for those years.
    a. Reported emissions should include net total emissions from processes involving CCS, including CCS applied to fossil energy and process-related emissions.
    b. By construction, emissions must be greater or equal to 0 in this section.
    c. To avoid double-counting, emissions accounted for in this section are not be accounted for in section 2) and vice versa.
2) Quantiles of the net emissions (CO_2 only) from the **combined sum** of:
    a. Agriculture, Forestry, and Other Land Use (AFOLU, $GtCO_2$)
    b. Sequestered emissions from Direct Air Capture (DAC, $GtCO_2$) and Bioenergy with Carbon Capture and Storage (BECCS, $GtCO_2$).
    c. By construction, total net emissions from this section may be positive (net CO_2 source) or negative (net CO_2 sink) in this section.
3) Quantiles of global CH_4 emissions ($GtCH_4$), including emissions from AFOLU.
4) Quantiles of global N_2O emissions (GtN_2O), including emissions from AFOLU

For each of the quantiles specified:

1) Quantile values will be linearly interpolated in time between each of the years elicited. Consequently, experts are specifying quantiles of piece-wise linear, non-overlapping trajectories.
    a. For example, the 5th percentile trajectory of emissions intensity represents a linear interpolation in time of the specified 5th percentiles for 2050, 2100, 2150, 2200, and 2300.

2) Quantiles for AFOLU, DAC, BECCS, CH_4, and N_2O by default will apply to all economic growth trajectories. Experts will, at their discretion, be able to provide such projections conditioned on economic growth in the same manner as for CO_2.

Additional Detail

The NAS recommended that, in order to provide for the long residence time of carbon dioxide in the atmosphere, the SCC model runs (and the socioeconomic projections required) should extend far enough into the future to account for the vast majority of resulting damages, and that the year 2300 was sufficient to meet this recommendation. The NAS further recommended that the appropriate way to address challenges inherent in quantifying the uncertainty for projections for such an extended time horizon is to use statistical methods in concert with results based upon the formal elicitation of experts, referred to as structured expert judgment.

RFF's program of research, under which this structured expert judgment on future emissions falls, is implementing these recommendations to yield a set of very long-run (multi-century) central projections for economic growth, population, and global emissions, with associated uncertainty bounds. The research employs statistical methods and expert elicitation to generate probability density functions for projections of each term of the following form of the IPAT identity:

*Emissions ≡ Population * (GDP/Population) * (Emissions/GDP)*

Previous work under the initiative has yielded country level PDFs of population and GDP/capita. Country-level PDFs of population to 2300 are available based upon Raftery and Ševčíková (under review, 2021). Country-level PDFs for GDP/capita to 2300 are available based upon the methodology of **Mueller, Stock, and Watson** (2020) used in concert with results from the *RFF Economic Growth Survey*. RFF's *Future Emissions Survey* is designed to provide PDFs of global emissions and related variables, conditioned on future economic growth, based upon an expert elicitation of ~12 experts on global emissions.

Structured Expert Judgment Methodology

The goal of structured expert judgment is to generate a probability distribution for one or more variables of interest by combining a set of individual distributions for the variables that have been provided by a set of experts. There are a number of ways to generate the resulting combined distribution, the simplest being to combine the set of individual distributions in equal weights. An alternate method, which generally provides advantages of narrower overall uncertainty distributions with greater statistical accuracy and has been shown to provide greater performance both **in-sample** and **out-of-sample,** performance-weights the experts according to their ability to quantify their uncertainty for a set of calibration variables from their field for which the true values are known. This latter approach is exemplified by the **Classical Model** for structured expert judgment, so called for its analogy with classical hypothesis testing. In its publication of its research on this topic, RFF will provide distributions generated based upon both approaches.

In the Classical Model, each of the experts on a panel quantifies his or her uncertainty with regard to variables of interest as well as with regard to a set of calibration variables from the subject area for which true values are known. Experts are treated as statistical hypotheses and scored on two performance metrics -- *statistical accuracy* and *informativeness* -- on the calibration variables.

Statistical accuracy: Roughly, an expert is statistically accurate if, in a statistical sense, 5% of the true values fall below his/her 5th percentiles, 45% of the realizations fall between his/her 50th and 5th percentiles, etc. More formally, the statistical accuracy of a given expert is measured as the probability (P-Value) of falsely rejecting the hypothesis that an expert's observed inter-percentile frequencies comply with his/her probabilistic assessments.

Informativeness: Informativeness is measured per variable as the Shannon relative information in an expert's distribution relative to a background measure. The background measure is (log) uniform on a 10% extension of the smallest interval containing all expert quantiles for the given variable. An expert's informativeness score is the average informativeness over all variables.

RFF Future Emissions Survey

Part 1: Practice questions

- *For each of the following values requested, please provide your 5th, 50th, and 95th percentiles. In other words: for the 5th percentile, provide the value for which the true value has a 1 in 20 chance of being less than; for the 50th percentile provide the value for which the true value has an equal chance of being greater or less than; and for the 95th percentile provide the value for which the true value has a 1 in 20 chance of exceeding.*
- *Please refrain from consulting outside sources, including those found on the internet, in answering these questions.*

The geometric mean of annual growth rates yields the rate at which constant growth over the full period would result in the observed changes in GDP levels from the beginning of the period to the end – e.g.: $GDP_{2017} = GDP_{1980} * ((1 + \text{geometric mean})^{38})$.

The geometric mean of the **global** annual growth rate of GDP from 1980 to 2017 was 2.86% per year.

3. **What is the geometric mean of the annual growth rate from 1980 to 2017 for Saudi Arabia?**

Also for Saudi Arabia, we are interested in your assessment of the **variability** of annual economic growth. One way of measuring such variability is by evaluating the Mean Absolute Deviation (MAD) with respect to the region's arithmetic mean for the full period.

$$MAD = \frac{1}{38} \sum_{y=1980}^{2017} |region\ growth\ rate\ in\ year\ y - \ mean\ region\ growth\ rate\ |$$

4. **What is the MAD of annual GDP growth for Saudi Arabia from 1980 to 2017?**

Part 2: Performance calibration questions

- *For each of the following values requested, please provide your 5th, 50th, and 95th percentiles. In other words: for the 5th percentile, provide the value for which the true value has a 1 in 20 chance of being less than; for the 50th percentile provide the value for which the true value has an equal chance of being greater or less than; and for the 95th percentile provide the value for which the true value has a 1 in 20 chance of exceeding.*
- *Please refrain from consulting outside sources, including those found on the internet, in answering these questions. For definitions see the Appendix.*

For the period **1960-2020** the **world** growth rate of GDP per capita in non-recession years was 2.1. In the five recession years (1975, 1982, 1991, 2009, 2020) the world growth average was -2.3.

The ***spread*** of world average growth between the non-recession years and recession years is 2.1 - (-2.3) = 4.4.

1. **What is the spread of average growth rate between non-recession years and world recession years for Advanced Economies?**

2. **What is the spread of average growth rate between non-recession years and world recession years for Emerging Markets and Developing Economies?**

3. **What is the spread of average growth rate between non-recession years and world recession years for East Asia and Pacific Countries?**

CO_2 emissions intensity is defined as CO_2 emissions (kg) per unit of GDP (2017$ PPP)

For the period 1990-2018, the **world percentage change** in CO_2 emissions intensity:

$$\frac{emissions\ intensity_{2018} - emissions\ intensity_{1990}}{emissions\ intensity_{1990}}$$

was **-0.331.**

4. **For the period 1990-2018, what was the percentage change in CO_2 emissions intensity for OECD members?**

5. **For the period 1990-2018, what was the percentage change in CO_2 emissions intensity for Sub-Saharan Africa?**

For the period 1990-2018, the year-over-year **world** percentage change in *CO_2 emissions intensity* was negative for 26 out of 28 years.

6. **For the period 1990-2018, how many years (out of 28 possible) was the year-over-year change in CO_2 emissions intensity negative for South Asia?**

7. **For the period 1990-2018, how many years (out of 28 possible) was the year-over-year change in CO_2 emissions intensity negative for Sub-Saharan Africa?**

Energy intensity is defined as energy consumed (ktoe) per GDP (2015$ PPP).

For the period 1990-2018, the **world percentage change** in *energy intensity*:

$$\frac{energy\ intensity_{2018} - energy\ intensity_{1990}}{energy\ intensity_{1990}}$$

was **-0.35.**

8. **For the period 1990-2018, what was the percentage change in energy intensity for the Middle East?**

9. **For the period 1990-2018, what was the percentage change in energy intensity for China?**

For the period 1980-2018, the percentage of primary energy for the **world** coming from renewable sources -- defined here as hydropower, solar, wind, geothermal, wave, tidal, and modern biofuels, but excluding traditional biomass – increased from 6.37% to 10.96%, for an absolute change of +4.59pp.

10. **For the period 1980-2018, what was the absolute change (percentage points) in primary energy coming from renewable sources for Venezuela?**

11. **For the period 1980-2018, what was the absolute change (percentage points) in primary energy coming from renewable sources for Portugal?**

Part 3: Elicitation

We will now ask you to provide quantiles of uncertainty (minimum, 5th, 50th, 95th, maximum, as well as others at your discretion) for several variables for the following two cases:

- **Evolving Policies**: Incorporating expected changes in technology, fuel use, and other conditions, *consistent with your expected evolution of future policy*.
- **Current Laws and Regulations**: Incorporating expected changes in technology, fuel use, and other conditions, *consistent with current on-the-books policies*.
    - Emissions distributions offered in under this case should represent current legislation and environmental regulations, including recent government actions for which implementing regulations were available as of August 2, 2021. The potential effects of proposed legislation, regulations, and standards—or sections of legislation that have been enacted but require funds to execute or do not have the required implementing regulations in place—should not be reflected here.

For each case, you will be asked to provide quantiles of uncertainty for the following for the years 2050, 2100, 2150, 2200, 2300:

1) Global CO_2 emissions ($GtCO_2$) for 3 future trajectories of GDP per capita representing the 2.5th percentile, 50th percentile, and 97.5th percentile of projected economic growth (three separate responses). You will also specify distributions for the minimum and maximum GDP per capita for those years.
    a. Reported emissions should include net total emissions from processes involving CCS, including CCS applied to fossil energy and process-related emissions.
    b. By construction, emissions must be greater or equal to 0 in this section.
    c. To avoid double-counting, emissions accounted for in this section are not be accounted for in section 2) and vice versa.

2) Quantiles of the net emissions (CO_2 only) from the **combined sum** of:
    a. Agriculture, Forestry, and Other Land Use (AFOLU, $GtCO_2$)
    b. Sequestered emissions from Direct Air Capture (DAC, $GtCO_2$) and Bioenergy with Carbon Capture and Storage (BECCS, $GtCO_2$).
    c. By construction, total net emissions from this section may be positive (net CO_2 source) or negative (net CO_2 sink) in this section.

3) Quantiles of global CH_4 emissions ($GtCH_4$), including emissions from AFOLU.
4) Quantiles of global N_2O emissions (GtN_2O), including emissions from AFOLU.

For each of the quantiles specified:

- Quantile values will be linearly interpolated in time between each of the years elicited. Consequently, experts are specifying quantiles of piece-wise linear, non-overlapping trajectories.
    - For example, the 5th percentile trajectory of emissions intensity represents a linear interpolation in time of the specified 5th percentiles for 2050, 2100, 2150, 2200, and 2300.

- Quantiles for AFOLU, DAC, BECCS, CH_4, and N_2O by default will apply to all economic growth trajectories. You will, at your discretion, be able to provide such projections conditioned on economic growth in the same manner as provided for CO_2.

- Please describe your rationale for the quantiles. For example, we will ask you to describe future narratives that would plausibly yield the reported levels of emissions with the likelihoods indicated.

- In order to capture your rationale, we will be taking notes throughout the elicitation. To facilitate such notetaking and ensure its veracity, we will also request that this section be recorded, solely for our internal use in processing the results. *Such recording is completely optional and at the discretion of the individual expert.*

- *Consulting outside sources for this part of the elicitation is permitted.*

Appendix A – Country classifications

Advanced Economies

Australia
Austria
Belgium
Canada
Cyprus
Czech Republic
Denmark
Estonia
Finland
France
Germany
Greece
Hong Kong SAR
China
Iceland
Ireland
Israel
Italy
Japan
Korea, Rep.
Latvia
Lithuania
Luxembourg
Malta
Netherlands
New Zealand
Norway
Portugal
Singapore
Slovak, Rep.
Slovenia
Spain
Sweden
Switzerland
United Kingdom
United States

Emerging Markets and Developing Economies

Albania*
Algeria*
Angola*
Argentina
Armenia
Azerbaijan*
Bahrain*
Belize
Benin
Bolivia*
Botswana
Brazil
Burkina Faso
Burundi
Cameroon*
Chad*
Chile
Colombia*
Congo, Dem. Rep.
Congo, Rep.*
Costa Rica
Côte d’Ivoire
Ecuador*
Equatorial Guinea*
Ethiopia
Gabon*
Gambia,
Ghana*
Lao PDR
Liberia
Madagascar
Malawi
Malaysia*
Mali
Mauritania
Mongolia
Morocco
Mozambique
Myanmar*
Namibia
Nicaragua
Niger
Nigeria*
Oman*
Papua New Guinea
Paraguay
Peru
Qatar*
Russia*
Rwanda
Saudi Arabia*
Senegal
Sierra Leone
South Africa
The Sudan*
Suriname
Afghanistan
Antigua and Barbuda
Bahamas, The
Bangladesh
Barbados
Belarus
Bhutan
Bosnia and Herzegovina
Bulgaria
Cabo Verde
Cambodia
China
Comoros
Croatia
Djibouti
Dominica
Dominican Republic
Egypt
El Salvador
Eritrea
Eswatini
Fiji
Georgia
Grenada
Haiti
Hungary
India
Jamaica
Pakistan
Palau
Panama
Philippines
Poland
Romania
Samoa
Serbia
Seychelles
Solomon Islands
Sri Lanka
St. Kitts and Nevis
St. Lucia
St. Vincent and the Grenadines
Thailand
Tonga
Tunisia
Turkey
Tuvalu
Vanuatu
Vietnam

Guatemala
Guinea
Guinea-Bissau
Guyana
Honduras
Indonesia*
Iran*
Iraq*
Kazakhstan*
Kenya
Kosovo
Kuwait*
Kyrgyz Republic
Tajikistan
Tanzania
Timor-Leste*
Togo
Turkmenistan*
Uganda
Ukraine
United Arab Emirates*
Uruguay
Uzbekistan
West Bank and Gaza
Zambia
Zimbabwe
Jordan
Kiribati
Lebanon
Lesotho
Maldives
Marshall Islands
Mauritius
Mexico
Micronesia, Fed. Sts.
Moldova, Rep.
Montenegro
Nepal
North Macedonia

East Asia and Pacific Countries

Cambodia
China
Fiji
Indonesia
Lao PDR
Malaysia
Mongolia
Myanmar
Papua New Guinea
Philippines
Solomon Islands
Thailand
Timor-Leste
Vietnam

OECD Countries

Australia
Austria
Belgium
Canada
Chile
Colombia
Costa Rica
Czech Republic
Denmark
Estonia
Finland
France
Germany
Greece
Hungary
Iceland
Ireland
Israel
Italy
Japan
Korea, Rep.
Latvia
Lithuania
Luxembourg
Mexico
Netherlands
New Zealand
Norway
Poland
Portugal
Slovak Republic
Slovenia
Spain
Sweden
Switzerland
Turkey
United Kingdom
United States

Sub Saharan Africa

Angola
Benin
Botswana
Burkina Faso
Burundi
Cabo Verde
Cameroon
Congo, Rep.
Cote D'Ivoire
Equatorial Guinea
Eritrea
Eswatini
Ethiopia
Gabon
Kenya
Lesotho
Liberia
Madagascar
Malawi
Mali
Mauritania
Nigeria
Rwanda
Sao Tome and Principe
Senegal
Seychelles
Sierra Leone
Somalia
Uganda
Zambia
Zimbabwe

Central African Republic	Gambia, The	Mauritius	South Africa
Chad	Ghana	Mozambique	South Sudan
Comoros	Guinea	Namibia	Tanzania
Congo, Dem. Rep	Guinea-Bissau	Niger	Togo

South Asia

Afghanistan	Sri Lanka
Bangladesh	Maldives
Bhutan	Nepal
India	Pakistan

Middle East

Islamic Republic of Iran	Iraq	Oman	Jordan
Saudi Arabia	Qatar	Bahrain	Lebanon
United Arab Emirates	Kuwait	Syria	Yemen