International Journal of Forecasting 28 (2012) 695–711

Contents lists available at SciVerse ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

The illusion of predictability: How regression statistics mislead experts

Emre Soyer[a], Robin M. Hogarth[a,b,*]

[a] Universitat Pompeu Fabra, Department of Economics & Business, Ramon Trias Fargas 25-27, 08005, Barcelona, Spain

[b] ICREA, Barcelona, Spain

ARTICLE INFO

Keywords:
Regression
Presentation formats
Probabilistic inference
Prediction
Graphics
Uncertainty

ABSTRACT

Does the manner in which results are presented in empirical studies affect perceptions of the predictability of the outcomes? Noting the predominant role of linear regression analysis in empirical economics, we asked 257 academic economists to make probabilistic inferences based on different presentations of the outputs of this statistical tool. The questions concerned the distribution of the dependent variable, conditional on known values of the independent variable. The answers based on the presentation mode that is standard in the literature demonstrated an illusion of predictability; the outcomes were perceived to be more predictable than could be justified by the model. In particular, many respondents failed to take the error term into account. Adding graphs did not improve the inference. Paradoxically, the respondents were more accurate when *only* graphs were provided (i.e., no regression statistics). The implications of our study suggest, inter alia, the need to reconsider the way in which empirical results are presented, and the possible provision of easy-to-use simulation tools that would enable readers of empirical papers to make accurate inferences.

1. Introduction

Much academic research in empirical economics involves determining whether or not one or several variables have causal effects on another variable. The statistical tool used for making such affirmations is typically regression analysis, where the terms "independent" and "dependent" are used to distinguish cause(s) from outcomes. The results from most analyses consist of statements as to whether or not particular independent variables are "significant" in affecting outcomes (the dependent variable), and most discussions of the importance of such variables focus on the "average" effects on outcomes of possible changes in inputs.

However, if the analysis is used for prediction, emphasizing only statistically significant average effects results in an incomplete characterization of the relationship between the independent and dependent variables. It is also essential to acknowledge the level of uncertainty inherent in outcomes of the dependent variable, conditional on values of the independent variable. For example, consider a decision maker who is pondering which actions to take and how much to do so in order to reach a certain goal. This requires conjectures to be formed about the individual outcomes that would result from specific inputs. Moreover, the answers to these questions depend not only on estimating average effects, but also on the distribution of possible effects around the average.

In this paper, we argue that the emphasis placed on determining average causal effects in the economics literature limits our ability to make correct probabilistic forecasts. In particular, the way in which results are presented in regression analyses obfuscates the uncertainty inherent in the dependent variable. As a consequence, consumers of the economic literature can be subject to what we call the "illusion of predictability".

Whereas it can be argued that the *way in which* information is presented should not affect rational interpretation and analysis, there is abundant psychological

* Corresponding author at: Universitat Pompeu Fabra, Department of Economics & Business, Ramon Trias Fargas 25-27, 08005, Barcelona, Spain. Tel.: +34 93 5422561; fax: +34 93 5421746.

E-mail addresses: emre.soyer@upf.edu (E. Soyer), robin.hogarth@upf.edu (R.M. Hogarth).

0169-2070/$ – see front matter © 2012 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.ijforecast.2012.02.002

evidence demonstrating that such presentation effects do occur. Many studies have shown, for example, the way in which subtle changes in questions designed to elicit preferences are subject to contextual influences (see, e.g., Kahneman & Tversky, 1979). Moreover, these have been reported in both controlled laboratory conditions and field studies involving appropriately motivated experts (Camerer, 2000; Thaler & Sunstein, 2008). The human information processing capacity is limited, and the manner in which attention is allocated has important implications for both revealed preferences and inferences (Simon, 1978).

Recently, Gigerenzer and his colleagues (Gigerenzer, Gaissmaier, Kurz-Milcke, Schwartz, & Woloshin, 2007) reviewed research on how probabilities and statistical information are presented, and consequently perceived, by individuals or specific groups that use them frequently in their decisions. They show that mistakes in probabilistic reasoning and the miscommunication of statistical information are common. Their work focuses mainly on the fields of medicine and law, where doctors, lawyers and judges fail to communicate crucial statistical information appropriately in particular situations, thereby leading to biased judgments that have a negative impact on others. One such example is the failure of gynecologists to infer the probability of cancer correctly, given the way in which mammography results are communicated.

We examine the way in which economists communicate statistical information. Specifically, we note that much of the work in empirical economics involves the estimation of average causal effects through the technique of regression analysis. However, when we asked a large sample of economists to use the standard reported outputs of the simplest form of regression analysis to make probabilistic forecasts for decision making purposes, nearly 70% of them experienced difficulties. The reason for this, we believe, is that current reporting practices focus attention on the uncertainty surrounding the model parameter estimates, and fail to highlight the uncertainty concerning outcomes of the dependent variable conditional on the model identified. On the other hand, when attention was directed appropriately – by graphical as opposed to tabular means – over 90% of our respondents made accurate inferences.

In the next section, we provide some background on the practice and evolution of reporting empirical results in economics journals. In Section 3 we provide information concerning the survey we conducted with economists, which involved them answering four decision-oriented questions based on a standard format for reporting the results of regression analyses. We employed six different conditions designed to assess the differential effects due to model fit (R^2) and different forms of graphical presentation (with and without accompanying statistics). In Section 4, we present our results. In brief, our study shows that the typical presentation format of econometric models and results – one based mainly on regression coefficients and their standard errors – leads economists to ignore the level of predictive uncertainty implied by the model and captured by the standard deviation of the estimated residuals. As a consequence, there is a considerable illusion of predictability. Adding graphs to the standard presentation of coefficients and standard errors does little to improve inferences. However, presenting the results in graphical fashion alone improved the accuracy. The implications of our findings, including suggested ways of improving statistical reporting, are discussed in Section 5.

2. Current practice

There are many sources of empirical analyses in economics. In order to obtain a representative sample of current practice, we selected all of the articles published in the 3rd issues (of each year) of four leading journals between 1998 and 2007 (441 articles). The journals were *American Economic Review* (AER), *Quarterly Journal of Economics* (QJE), *Review of Economic Studies* (RES) and *Journal of Political Economy* (JPE). Among these articles, we excluded those with time series analyses, and only included those with cross-sectional analyses where the authors identify one or more independent variables as statistically significant causes of relevant economic and social outcomes. Our aim is to determine how the consumers of this literature translate the findings about average causal effects into perceptions of predictability.

Many of the articles published in these journals are empirical. Over 70% of the empirical analyses use variations of regression analysis, of which 75% have linear specifications. Regression analysis is clearly the most prominent tool used by economists to test hypotheses and identify relationships among economic and social variables.

In economics journals, empirical studies follow a common procedure for displaying and evaluating results. Typically, authors provide a table that displays the descriptive statistics of the sample used in the analysis. Either before or after this display, they describe the specification of the model on which the analysis is based, then provide the regression results in detailed tables. In most cases, these results include the coefficient estimates and their standard errors, along with other frequently reported statistics, such as the number of observations and the R^2 values.

Table 1 summarizes these details for the sample of studies referred to above. It shows that, apart from the regression coefficients and their standard errors (or t-statistics), there is not much agreement as to what else should be reported. The data therefore suggest that economists probably understand the inferences that can be made about regression coefficients or the *average* impact of manipulating an independent variable quite well; however, their ability to make inferences about other probabilistic implications may be less well developed (e.g., predicting individual outcomes conditional on specific inputs).

It is not clear when, how, or why the above manner of presenting regression results in publications emerged. No procedure is ever explicitly stated in the submission guidelines for the highly ranked journals. Moreover, popular econometric textbooks, such as those of Greene (2003), Gujarati and Porter (2009) and Judge, Griffiths, Hill, and Lee (1985) do not explain specifically how to present results or how to use them for decision making. Hendry and Nielsen (2007) address issues regarding prediction in more detail than other similar textbooks. Another

Table 1
Distribution of types of statistics provided by studies in our sample of economics journals.

Studies that:	Journals					% of total
	AER	QJE	JPE	RES	Total	
…use linear regression analysis	42	41	15	13	111	*x*
…provide both the sample standard deviation of the dependent variable(s) and the R^2 statistic	16	27	11	12	66	59
…provide R^2 statistics	30	32	15	12	89	80
…provide the sample standard deviation of the dependent variable(s)	21	32	11	13	77	69
…provide the estimated constant, along with its standard error	19	14	4	1	38	34
…provide a scatter plot	19	16	5	2	42	38
…provide the standard error of the regression (SER)	5	3	1	1	10	9

exception is Wooldridge (2008), who dedicates several sections to presentation issues. His outline suggests that a good summary consists of a table with selected coefficient estimates and their standard errors, R^2 statistics, a constant, and the numbers of observations. Indeed, this is consistent with today's practice. More than 60% of the articles in Table 1 follow a similar procedure.

Zellner (1984) conducted a survey of statistical practice based on articles published in 1978 in the AER, JPE, *International Economic Review*, *Journal of Econometrics* and *Econometrica*. He documented confusion as to the meaning of tests of significance, and proposed Bayesian methods for overcoming theoretical and practical problems. Similarly, McCloskey and Ziliak (1996) provided an illuminating study of statistical practice based on articles published in AER in the 1980s. They demonstrated that there was widespread confusion in the interpretation of statistical results, due to a confounding of the concepts of statistical and economic or substantive significance. Too many results depended on whether the *t*- or other statistics exceeded arbitrarily defined limits. In follow-up studies, Ziliak and McCloskey (2004, 2008) report that, if anything, this situation worsened in the 1990s (see also Zellner, 2004).

Empirical finance has developed an illuminating way of determining the significance of findings. In this field, once statistical analysis has identified a variable as being "important" in affecting, say, stock returns, it is standard to assess "how important" it is by evaluating the performance of simulated stock portfolios that use the variable (see, e.g., Carhart, 1997, and Jensen, 1968).

In psychology, augmenting significance tests with the effect size became common practice in the 1980s. For example, in its submission guidelines, *Psychological Science*, the flagship journal of the Association for Psychological Science, explicitly states, "effect sizes should accompany major results. When relevant, bar and line graphs should include distributional information, usually confidence intervals or standard errors of the mean".

In forecasting, Armstrong (2007) initiated a discussion on not only the necessity of using effect size measures when identifying relationships among variables, but also the fact that significance tests should be avoided when doing so. He argues that the results of significance tests are often misinterpreted, and even when presented and interpreted correctly, they do not contribute to the decision making process. Schwab and Starbuck (2009) make an analogous argument for management science.

In interpreting the results of linear regression analysis from a decision making and predictive perspective, two statistics can convey a message to readers about the level of uncertainty in the results. These are R^2 and the Standard Error of the Regression (SER).[1] As a bounded and standardized quantity, R^2 describes the fit of a model. SER, on the other hand, provides information on the degree of predictability in the metric of the dependent variable.

Table 1 shows that SER is practically never given in the presentation of results: less than 10% of the studies with linear specifications provide it. R^2 is the prevalent statistic reported to give an indication of model fit. This is the case for 80% of published articles with a linear specification. Table 1 also shows that more than 40% of the publications in our sample that utilize a linear regression analysis (excluding studies that base their main results on an IV regression) provide no information on either R^2 or the standard deviation of the dependent variable. Hence, a decision maker consulting the results of these studies cannot infer much about either the unexplained variance within the dependent variable or the cloud of data points to which the regression line is fitted. Alternatively, a scatter plot would be essential in order to indicate the degree of uncertainty. However, less than 40% of the publications in our sample provide a graph with actual observations.

Given the prevalence of empirical analyses and their potential use for decision making and prediction, debates about how to present results are important. However, it is important that such debates be informed by evidence as to the way in which knowledgeable individuals use currently available tools for making probabilistic inferences, and the way in which different presentation formats affect judgment. Our goal is to provide such evidence.

3. The survey

3.1. Goal and design

How do knowledgeable individuals (economists) interpret specific decision making implications of the standard output of a regression analysis? To find out, we used the following criteria to select the survey questions. First, we provided information about a well-specified model that strictly met the underlying assumptions of linear regression analysis. Second, the model was straightforward, in

[1] Some sources refer to SER as the Standard Error of Estimates, or SEE (see RATS), while others refer to it as the root Mean Squared Error or root-MSE (see STATA). Wooldridge (2008) uses the term Standard Error of the Regression (SER), defining it as "an estimator of the standard deviation of the error term".

that it had only one independent variable. Third, all of the information necessary for solving the problems posed was available from the output provided. Fourth, although sufficient information was available, respondents had to apply knowledge about statistical inference in order to make the calculations necessary for answering the questions.

This last criterion is the most demanding, because whereas economists may be used to interpreting the statistical significance of regression coefficients, they typically do not assess the uncertainties involved in prediction when an independent variable is changed or manipulated (apart from making "on average" statements that give no hint as to the distribution around the average).

Our study required respondents to answer four decision making questions, after being provided with information about a correctly specified regression analysis. There were six different conditions, which varied in the overall fit of the regression model (Conditions 1, 3, and 5 with $R^2 = 0.50$, the others with $R^2 = 0.25$), as well as in the amount and type of information provided. Figs. 1 and 2 report the information provided to the respondents for Conditions 1 and 2, which is similar in form and content to the outputs of many reports in the economic literature (and consistent with Wooldridge, 2008). Conditions 3 and 4 used the same tables, but provided the bivariate scatter-plots of the dependent and independent variables in addition to the standard deviation of the estimated residuals—see Figs. 3 and 4. In Conditions 5 and 6, the statistical outputs of the regression analyses were not provided, but the bivariate graphs of the dependent and independent variables were, as in Figs. 3 and 4.[2] In other words, for these two conditions we were intrigued by what would happen if respondents were limited to only consulting graphs.

Similarly to our survey on current practice in Section 2, we again restrict our attention to cross-sectional analyses in our experimental conditions. We are primarily concerned with determining the way in which findings on average causal effects are used for predictions and decision making. Our variations over different conditions would not be valid for time series studies, where the R^2 statistic does not provide information on the model fit. It is important to add that results are also discussed in the text in published papers. These discussions, which are mostly confined to certain coefficient estimates and their statistical significance levels, might distract decision makers from the uncertainties about outcomes. None of our conditions involve such discussions.

3.2. Questions

For Conditions 1, 3, and 5, we asked the following questions:

1. What would be the minimum value of X that an individual would need to make sure that s/he obtains a positive outcome ($Y > 0$) with 95% probability?
2. What minimum, positive value of X would make sure, with 95% probability, that the individual obtains more Y than a person who has $X = 0$?
3. Given that the 95% confidence interval for β is (0.936, 1.067), if an individual has $X = 1$, what would be the probability that s/he gets $Y > 0.936$?
4. If an individual has $X = 1$, what would be the probability that s/he gets $Y > 1.001$ (i.e. the point estimate)?

The questions for Conditions 2, 4, and 6 were the same, except that the confidence interval for β is (0.911, 1.130), and we ask about the probabilities of obtaining $Y > 0.911$ and $Y > 1.02$, given $X = 1$, in questions 3 and 4 respectively. All four questions are reasonable, in that they seek answers to questions that would be of interest to decision makers. However, they are not the types of questions that reports in economics journals usually lead readers to pose, and thus, they test a respondent's ability to reason in a correct statistical manner given the information provided. In Appendix A, we provide the rationale behind the questions and the correct answers.

3.3. Respondents and method

We sent web-based surveys to faculty members in economics departments at leading universities worldwide. From the top 150 departments, ranked by numbers of econometric publications between 1989 and 2005 (Baltagi, 2007, Table 3), we randomly selected 113.[3] Within each department, we randomly selected up to 36 faculty members. We ordered them alphabetically by their names and assigned Condition 1 to the first person, Condition 2 to the second person,..., Condition 6 to the sixth person, then again Condition 1 to the seventh person, and so on.

We conducted the survey online by personally sending a link for the survey, along with a short explanation, to the professional email address of each prospective participant. In this way, we managed to keep the survey strictly anonymous. We do know the large pool of institutions to which the participants belong, but have no means of identifying the individual sources of the answers. The participants answered the survey voluntarily. They had no time constraints and were allowed to use calculators or computers if they wished. We told all prospective participants that, at the completion of the research, the study along with the feedback on questions and answers would be posted on the web and that they would be notified,[4] but did not offer them any economic incentives for participation.

As can be seen from Table 2, we dispatched a total of 3013 requests to participate. About one-quarter of potential respondents (26%) opened the survey and, we presume,

2 We thank Rosemarie Nagel for suggesting that we include Conditions 5 and 6.

3 We stopped sampling universities once we had at least 30 individual responses for each question asked. A few universities were not included in our sample because their webpages did not facilitate access to potential respondents. This was more frequent for non-US universities. For reasons of confidentiality, we do not identify any of these universities.

4 In fact, this was done right after a first draft of the paper had been written.

Consider the econometric model

$$Y_i = C + \beta X_i + e_i$$

Where:

Y : Economic payoff, given the choice of X.
X : A continuous choice variable which is costly to undertake
C : Constant
β : The effect of X on Y
e : Random perturbation; $e_i \mid X_i \sim N[0, \sigma^2]$ with $E(e_i)=0$, $Cov(e_i, e_j)=0$ and $Cov(e_i, X_i)=0$.

In this setting, the goal is to estimate β and C, based on a random sample of X and Y with 1000 observations. The sample statistics are as follows:

Variable	Mean	Std. Dev.
X	50.72	28.12
Y	51.11	40.78

The OLS fit of the model to this sample gives the following results:

	Dependent Variable: *Y*
X	1.001 (0.033)**
Constant	0.32 (1.92)
R^2	0.50
N	1 000

Standard errors in parentheses
** Significant at 95% confidence level
N is the number of observations

Results indicate that constant C is not statistically different from zero and that X has a statistically significant positive effect on Y. β is estimated to be 1.001.

Suppose that this model is indeed a very good approximation of the real world relation between X and Y, and that the linear estimation is suitable. Furthermore, among alternative specifications, this model is the one that gives the highest R-squared.

The above result is a useful tool for decision-making purposes: It links the economic payoffs Y to the choice variable X. One can now use this relation to predict one's payoffs or to select their X and to obtain desired levels of Y. More importantly, the above model links Y and X correctly. This is crucial because increasing X is costly and knowing this true relationship helps individuals make more accurate decisions.

Fig. 1. Presentation of Condition 1. This mimics the methodology of 60% of the publications that were surveyed, and also the suggestions of Wooldridge (2008).

looked at the set-ups and questions. About a third of these (or 9% of all potential respondents) actually completed the survey. The proportions of potential respondents who opened the surveys and responded was highest for Conditions 5 and 6 (40%), as opposed to the 30% and 32% in Conditions 1 and 2, and 3 and 4, respectively. The average time taken to complete the survey was also lowest for Conditions 5 and 6 (see the notes to Table 2). We consider these outcomes again when we discuss the results below.

Table 2 documents characteristics of our respondents. In terms of position, the majority (59%) are at the rank of Associate Professor or higher. They also work in a wide variety of fields within the economics profession. Thirteen percent of respondents classified themselves as econometricians, and more than two-thirds (77%) used regression analysis in their work (41% "often" or "always").

4. Results

4.1. Condition 1

The respondents' answers to Condition 1 are summarized in Fig. 5. Three answers were removed from the data, being only "I don't know", or "?". For the first two questions, responses within ±5 of the correct amount were considered correct. For questions 3 and 4, we considered correct any responses that were within ±5% of the answer.

Variable	Mean	Std. Dev.
X	49.51	28.74
Y	51.22	59.25

	Dependent Variable: *Y*
X	1.02 (0.056)**
Constant	0.61 (3.74)
R^2	0.25
N	1 000

Notes:
Standard errors are in parentheses
** Significant at the 95% confidence level
N is the number of observations

Fig. 2. Tables in Condition 2. The rest of the presentation is the same as Fig. 1.

We also regarded as correct the responses of four participants who did not provide numerical estimates, but mentioned that the answer was related mainly to the error term and its variance (there were 21 such responses across all conditions). The questions and the correct answers are displayed in the titles of the histograms in Fig. 5.

Most answers to the first three questions were incorrect. They suggest that the presentation leads to the respondents only evaluating the results through the coefficient estimates, and obscures the uncertainty implicit in the dependent variable. Specifically, Fig. 5 shows that:

1. 72% of the participants believe that for an individual to obtain a positive outcome with 95% probability, a small X ($X < 10$) would be enough, given the regression results. The majority state that any small positive value of X would be sufficient to obtain a positive outcome with 95% probability. In actual fact, in order to obtain a positive outcome with 95% probability, a decision maker should choose approximately $X = 47$.
2. 71% of the answers to the second question suggest that for an individual to be better off than another person with $X = 0$, with 95% probability, a small value of X ($X < 10$) would be sufficient. In fact, given that the

Fig. 3. Bivariate scatter plot of Condition 1 and information on SER. Both were provided to participants in Condition 3, along with the estimation results. Only the graph was provided in Condition 5.

Fig. 4. Bivariate scatter plot of Condition 2 and information on SER. Both were provided to participants in Condition 4, along with estimation results. Only the graph was provided in Condition 6.

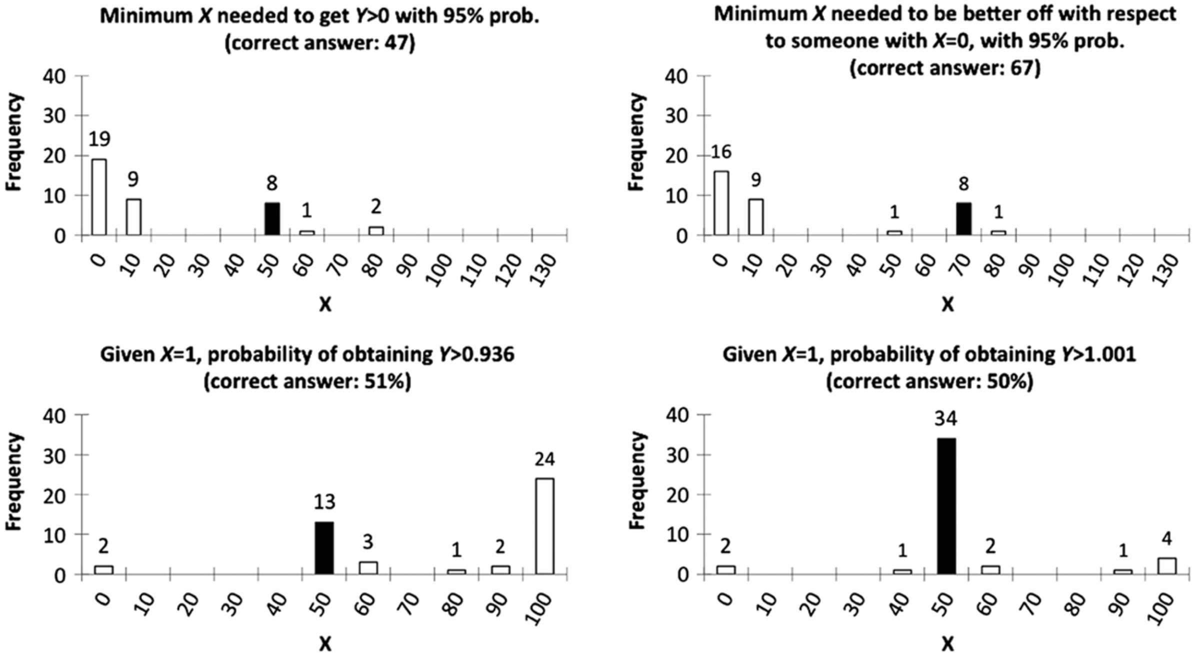

Fig. 5. Histograms for the responses to Condition 1. The top-left figure shows answers to question 1, the one on the top-right shows answers to question 2, the one on the bottom-left those to question 3, and the one on the bottom-right those to question 4. Each histogram also displays the question and the approximate correct answer. The dark column identifies the responses that we considered correct. Above each column is the number of participants who gave that particular answer. There were 39, 35, 45 and 44 responses to questions 1–4, respectively.

person with $X = 0$ will also be subject to a random shock, the value of X needed to ensure this condition is approximately 67.

3. 60% of the participants suggest that, given $X = 1$, the probability of obtaining an outcome that is above the lower bound of the estimated coefficient's 95% confidence interval is very high (greater than 80%). Instead, the correct probability is approximately 51%, as the uncertainty around the coefficient estimates in this case is small compared to the uncertainty due to the error term.
4. 84% of participants gave an approximately correct answer of 50% to question 4.

The participants' answers to the first two questions suggest that the uncertainty affecting Y is not directly visible in the presentation of the results. The answers to question 3, on the other hand, shed light on what the majority of our sample sees as being the main source of fluctuation in the dependent variable. The results suggest that it is the uncertainty concerning the estimated coefficients that is seen to be important, not the magnitude of the SER. In the jargon of popular econometrics texts, whereas respondents were sensitive to one of the two sources of prediction error, namely the sampling error, they ignored the error term of the regression equation. The apparent invisibility of the random component in the presentation lures respondents into disregarding the error term, and into confusing an outcome with its estimated expected value.

In their answers to questions 3 and 4, the majority of participants claim that if someone chooses $X = 1$, there is a 50% probability of obtaining $Y > 1.001$, but that obtaining $Y > 0.936$ is almost certain. Incidentally, the high rate of correct answers to question 4 suggests that the failure to respond accurately to questions 1–3 was not because participants failed to pay attention to the task (i.e., they were not responding "randomly").

Our findings echo those of Lawrence and Makridakis (1989), who showed in an experiment that decision makers tend to construct confidence intervals of forecasts using estimated coefficients, and fail to correctly take into account the randomness inherent in the process they are evaluating. Our results are also consistent with those of Goldstein and Taleb (2007), who showed that failing to interpret a statistic appropriately can lead to incorrect assessments of risk.

In summary, the results of Condition 1 show that the most common way of displaying results in the empirical economics literature leads to an illusion of predictability, in that part of the uncertainty is invisible to the respondents. In Condition 2, we test this interpretation by seeing whether the answers to Condition 1 are robust to different levels of uncertainty.

4.2. Conditions 2–4

If the presentation of the results causes the error term to be ignored, then the answers of the decision makers should not change in different set-ups, regardless of the variance of the error term, provided that its expectation is zero. To test this, we change only the variance of the error term in Condition 2 (see Fig. 2). Conditions 3 and 4 replicate Conditions 1 and 2, except that we add scatter plots and SER statistics — see Figs. 3 and 4.

Table 2
Characteristics of respondents.

Condition	1	2	3	4	5	6	Total	%
Requests to participate	568	531	548	510	438	418	3013	–
Requests opened	143	152	140	131	113	98	777	26
Surveys completed	45	45	49	38	36	44	257	9
Position								
Professor	17	14	19	18	17	22	107	42
Associate professor	8	7	12	10	6	2	45	18
Assistant professor	12	18	16	9	9	12	76	30
Lecturer	6	4	1	1	3	3	18	7
Other	2	2	1	0	1	5	11	4
Total	45	45	49	38	36	44	257	
Use of regression analysis								
Never	7	5	11	11	6	15	55	23
Some	11	16	17	10	17	13	84	36
Often	16	14	7	7	7	8	59	25
Always	5	5	8	6	6	7	37	16
Total	39	40	43	34	36	43	235	
Average minutes spent	11.6	10.3	7.4	7.5	5.7	6.5	8.1	
⟨Std. dev.⟩	⟨12.0⟩	⟨7.8⟩	⟨7.1⟩	⟨5.3⟩	⟨3.9⟩	⟨6.0⟩	⟨7.7⟩	

Table 3
Comparison of results for Conditions 1 to 6.

Condition	1	2	3	4	5	6
R^2	*0.50*	*0.25*	*0.50*	*0.25*	*0.50*	*0.25*
Scatter plot	*No*	*No*	*Yes*	*Yes*	*Yes*	*Yes*
Estimation resuls	*Yes*	*Yes*	*Yes*	*Yes*	*No*	*No*
Percentage of participants whose answer to:						
Question (1) was $X < 10$ (*Incorrect*)	**72**	**67**	**61**	**41**	**3**	**7**
Question (2) was $X < 10$ (*Incorrect*)	**71**	**70**	**67**	**47**	**3**	**15**
Question (3) was above 80% (*Incorrect*)	**60**	**64**	**63**	**32**	**9**	**7**
Question (4) was approx. 50% (*Correct*)	**84**	**88**	**76**	**84**	**91**	**93**
Approximate correct answers are						
Question 1	47	82	47	82	47	82
Question 2	67	116	67	116	67	116
Question 3 (%)	51	51	51	51	51	51
Question 4 (%)	50	50	50	50	50	50
Number of participants						
Question 1	39	36	44	32	31	41
Question 2	35	30	39	32	30	39
Question 3	45	42	49	37	32	43
Question 4	44	41	49	37	32	43

Notes:
Question (1) What would be the minimum value of X that an individual would need to make sure that s/he obtains a positive outcome ($Y > 0$) with 95% probability?
Question (2) What minimum, positive value of X would make sure, with 95% probability, that the individual obtains more Y than a person who has $X = 0$?
Question (3) Given that the 95% confidence interval for β is (a, b), if an individual has $X = 1$, what would be the probability that s/he gets $Y > a$?
Question (4) If an individual has $X = 1$, what would be the probability that s/he gets $Y > \hat{\beta}$?
In Conditions 1, 3 and 5, $a = 0.936$, $b = 1.067$ and $\hat{\beta} = 1.001$; in Conditions 2, 4 and 6, $a = 0.911$, $b = 1.13$ and $\hat{\beta} = 1.02$.

The histograms of the responses to the four questions in Conditions 2–4 are remarkably similar to those of Condition 1 (see Appendix B). These similarities are displayed in Table 3.

The similarities between the responses in Conditions 1 and 2 show that – under the influence of the current methodology – economists are led to overestimate the effects of explanatory factors on economic outcomes. The misperceptions demonstrated in the respondents' answers suggest that the way in which regression results are presented in publications can prevent even knowledgeable individuals from differentiating among different clouds of data points and uncertainties. At an early stage of our investigation, we also conducted the same survey (using Conditions 1 and 2) with a group of 50 graduate students in economics at Universitat Pompeu Fabra who had recently

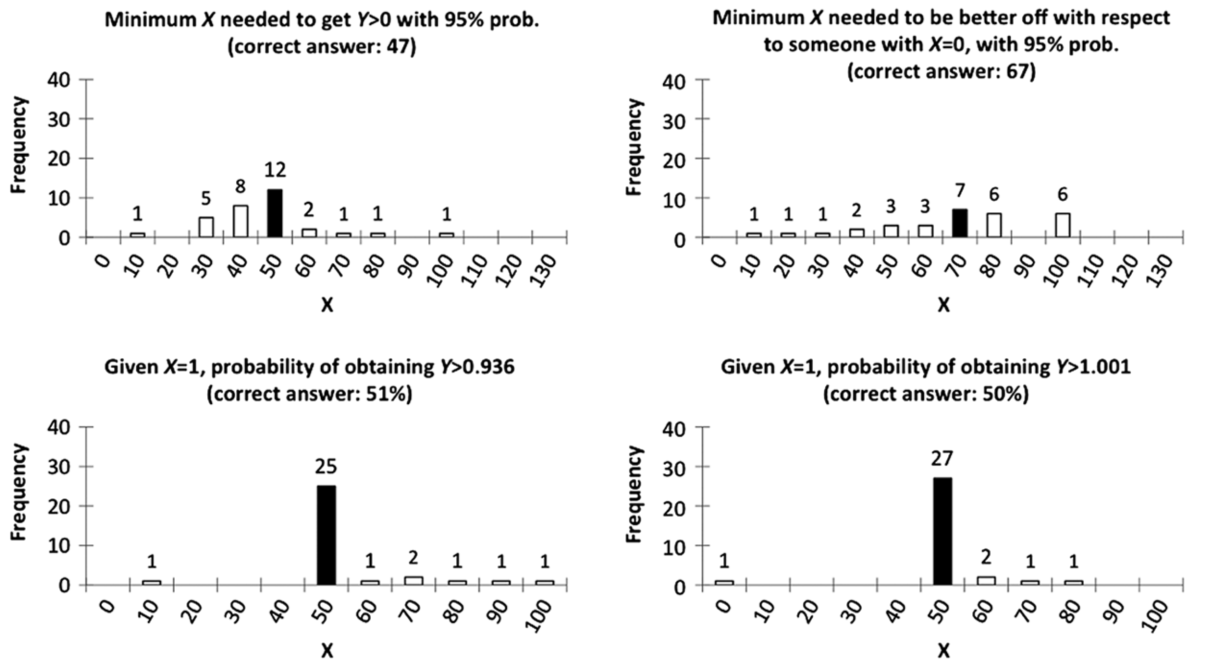

Fig. 6. Histograms for the responses to Condition 5. The top-left figure shows answers to question 1, the one on the top-right shows answers to question 2, the one on the bottom-left those to question 3, and the one on the bottom-right those to question 4. Each histogram also displays the question and the approximate correct answer. The dark column identifies the responses that we considered correct. Above each column is the number of participants who gave that particular answer. There were 31, 30, 32 and 32 responses to questions 1–4, respectively.

taken an advanced econometrics course, as well as with 30 academic social scientists (recruited through the European Association for Decision Making). The results (not reported here) were similar to those of our sample of economists, and suggest that the origins of the misperceptions can be traced back to the methodology, as opposed to professional backgrounds.

Table 3 indicates that when the representation is augmented with a graph of actual observations and with statistical information on the magnitude of the error term (SER), the perceptions of the relevant uncertainty, and consequently the predictions, improve. However, around half of the participants still fail to take the error term into account when making predictions, and give answers similar to those in Conditions 1 and 2 (see Appendix B for histograms of responses to Conditions 3 and 4). This suggests that respondents still rely mainly on the table showing the estimated coefficients and their standard errors as the main tool for assessing uncertainty. Since the information provided in Conditions 3 and 4 is rarely provided in published papers, this does not provide much hope for improvement. Possibly more drastic changes are necessary. Conditions 5 and 6 were designed to test this suggestion.

4.3. Conditions 5 and 6

Our results so far suggest that, when making predictions using regression analysis, economists pay an excessive amount of attention to coefficient estimates and their standard errors, and fail to consider the uncertainty inherent in the relationships between the dependent and independent variables. What happens, therefore, when they cannot see estimates of coefficients and related statistics, but have only a bivariate scatter plot? This is the essence of Conditions 5 and 6 (see the graphs in Figs. 3 and 4).

Fig. 6 displays the histograms for the responses to the four questions in Condition 5. The responses to Condition 6 were similar, and the histograms are displayed in Appendix B. These histograms show that the participants are much more accurate in their assessments of uncertainty now than in the previous conditions (see also Table 3). In fact, when the coefficient estimates are not available, they are forced to attend solely to the graph, which depicts the uncertainty within the dependent variable adequately. This further suggests that scant attention was paid to the graphs when the coefficient estimates were present. Despite the unrealistic manner of presenting the results, Conditions 5 and 6 show that a simple graph can be better suited to assessing the predictability of an outcome than a table with coefficient estimates, or even than a presentation that includes both a graph and a table.

In Conditions 5 and 6, most of the participants, including some of those who made the most accurate predictions, protested in their comments about the insufficiency of the information provided for the task. They claimed that it was impossible to determine the answers without the coefficient estimates, and that all they did was to "guess" the outcomes approximately. Yet their guesses were more accurate than the predictions from the previous conditions, which were the result of a careful investigation of the coefficient estimates and time-consuming computations. Indeed, as Table 2 indicates, the respondents in Conditions 5 and 6 spent significantly less time on the task than those in Conditions 1 and 2 ($t(40) = 2.71$ and $t(40) = 2.38$, $p = 0.01$ and 0.02, respectively).

4.4. Effects of training and experience

Table 2 shows that our sample of 257 economists varied widely in terms of professorial rank and the use of regression analysis in their work. We failed to find any relationship between the numbers of correct answers and either professorial rank or frequency of using regression analysis. A higher percentage of statisticians, financial economists and econometricians performed well relative to the average respondent (with 64%, 56%, and 51% providing correct answers, respectively, compared to the overall average of 35%). When the answers were more accurate, the average time spent was also slightly greater (10.2 min versus 9.3). Appendix C shows in detail the characteristics and proportions of respondents who gave accurate answers in Conditions 1–4.

5. Discussion

We conducted a survey of the probabilistic predictions made by economists on the basis of regression outputs similar to those published in leading economics journals. When given only the regression statistics which are typically reported in such journals, many respondents made inappropriate inferences. In particular, they seemed to locate the uncertainty of prediction in estimates of the regression coefficients, but not in the standard error of the regression (SER). Indeed, the responses hardly differed depending on whether the fit of the estimated model was 0.25 or 0.50.

We also provided some respondents with scatter plots of the regression, together with explicit information on the SER. However, this had only a small ameliorative effect, suggesting that respondents relied principally on the regression statistics (e.g., coefficients and their standard errors) when making their judgments. Finally, we forced other respondents to rely on a graphical representation by providing only a scatter plot, with no regression statistics. Members of this group complained that they did not have sufficient information, but – most importantly – were more accurate in their responses than the other groups, and also took less time to answer.

Several issues could be raised about our study, in relation to the nature of the questions asked, the specific respondents recruited, and their motivations for answering our questions. We now address these issues.

First, we deliberately asked questions that are usually not posed in journal articles because we sought to illuminate economists' appreciations of the predictability of economic relationships, as opposed to the assessment of the "significance" of certain variables (McCloskey & Ziliak, 1996; Ziliak & McCloskey, 2004, 2008). This is important. For example, even though economics articles do not typically address explicit decision making questions, the models can be used to estimate, say, the probability of reaching a given level of output for a specific level of input, as well as the economic significance of the findings. It is also important to understand that a policy that achieves a significantly positive effect "on average" might still be undesirable, because it leaves a large fraction of the population worse off. Hence, the questions are essential but "tricky" only in the sense that they are not the sorts of questions which economists typically ask.

Second, as was noted earlier, 26% of potential respondents took the time to open (and look at?) our survey questions, and 9% answered. Does this mean that our respondents were biased, and if so, in what direction were they biased? We clearly cannot answer this question, but we can state that our sample contained a substantial number of respondents (257), who represent various different characteristics of academic economists. Moreover, they were relevant respondents, in that they were recruited worldwide from leading departments of economics, as judged by publications in econometrics (Baltagi, 2007).

Third, by maintaining anonymity in the responses, we were unable to offer incentives to our respondents. However, would incentives have made a difference? Clearly, we cannot say without conducting a specific study. However, the consensus from previous results in experimental economics is that incentives increase effort and reduce the variance in the responses, but do not necessarily increase the average accuracy (Camerer & Hogarth, 1999). We also note that when professionals are asked questions which relate to their level of competence, there is little incentive to provide casual answers. Interestingly, our survey is a good simulation of the circumstances under which many economists read journal articles: there are no explicit monetary incentives; readers do not wish to make additional computations or to do work to fill in gaps left by the authors; and time is precious. Thus, the presentation of results is crucial.

Since our investigation concerns the way in which statistical results are presented in academic journals, it is important to ask what specific audience authors have in mind. The goal in leading economics journals is scientific: to identify which variables have an impact on some economic output and to assess the strength of the relationship. Indeed, the discussion of results often involves terms such as a "strong" effect, where the rhetoric reflects the size of t-statistics and the like. Moreover, the strength of a relationship is often described only from the perspective of an average effect, e.g., that a unit increase in an independent variable implies a δ increase in the dependent variable, on average.

As preliminary statements of the relevance of specific economic variables, this practice is acceptable. Indeed, although authors undoubtedly want to emphasize the scientific importance of their findings, we see no evidence of deliberate attempts to mislead readers into believing that the results imply a greater control over the dependent variable than is, in fact, the case. In addition, the papers have been reviewed by peers who are typically not shy about expressing their reservations. However, from a decision making perspective, the typical form of presentation can lead to an illusion of predictability of the outcomes, given the underlying regression model. Specifically, there can be a considerable degree of variability around the expectations of effects, which needs to be calibrated in the interpretation of results. Thus, readers who don't "go beyond the information given" and take the trouble to calculate, say, the implications of some decision-oriented questions, may gain an inaccurate view of the results obtained.

At one level, it could be argued that the principle of *caveat emptor* should apply. That is, consumers of economic research should know how to use the information provided, and it is their responsibility to assess the uncertainty appropriately. It is not the fault of either the authors or the journals if they cannot. However, we make two arguments against the *caveat emptor* principle, as applied here.

First, as has been demonstrated by our survey, even knowledgeable economists experience difficulty in going beyond the information provided in typical outputs of regression analysis. If one wants to make the argument that people "ought" to do something, then it should also be clearly demonstrated that they "can".

Second, given the vast numbers of economic reports available, it is unlikely that most readers will take the necessary steps to go beyond the information provided. As a consequence, by reading journals in economics they will necessarily acquire a false impression of what the knowledge gained from economic research allows one to say. In short, they will believe that economic outputs are far more predictable than is actually the case.

We make all of the above statements under the assumption that econometric models describe empirical phenomena appropriately. In reality, such models may suffer from a variety of problems associated with the omission of key variables, measurement errors, multicollinearity, or estimating the future values of predictors. It can only be shown that model assumptions are, at best, approximately satisfied (they are not "rejected" by the data). Moreover, whereas the model-data fit is maximized within the particular sample observed, there is no guarantee that the estimated relationships will be maintained in other samples. Indeed, the R^2 value estimated on a fitting sample inevitably "shrinks" when predicting to a new sample, and estimating the amount of shrinkage *a priori* is problematic. There is also evidence that statistical significance is often wrongly associated with replicability (Tversky & Kahneman, 1971; see also Hubbard & Armstrong, 1994). Possibly, if authors discussed these issues further, people's perceptions of the predictability of outcomes would improve. However, these considerations are beyond the scope of the present study.

Furthermore, because our aim was to isolate the impact of the presentation mode on predictions, we made many simplifying assumptions. For instance, errors that are heteroskedastic and non-normally distributed, or the presence of fewer observations at the more extreme values of the dependent variable would also increase prediction error. Even though many estimation procedures do not require assumptions, such as that of normally distributed random disturbances, in order to obtain consistent estimates, the explanations which they provide through coefficient estimates and average values would be less accurate if the law of large numbers did not hold. Hence, in more realistic scenarios, where our assumptions are not valid, decisions that are weighted towards expected values and coefficient estimates would be even less accurate than our results indicate.

How then can current practice be improved? Our results show that providing graphs alone led to the most accurate inferences. However, since this excludes the actual statistical analysis evaluating the relationships between different variables, we do not deem it a practical solution. Nevertheless, we do believe that it is appropriate to present graphs together with summary statistics, as we did in Conditions 3 and 4, although this methodology does not eliminate the problem.

We seriously doubt that any substantial modification of current practice will be accepted. We therefore suggest *augmenting* reports by requiring the authors to provide internet links to simulation tools. These could explore different implications of the analysis, as well as let readers pose different probabilistic questions. In short, we propose that tools be provided which allow readers to experience the uncertainty in the outcomes of the regression.[5]

In fact, we recently embarked on a test of the effectiveness of simulations in facilitating probabilistic inferences (Hogarth & Soyer, 2011). In two experiments, conducted with participants at varying levels of statistical sophistication, respondents were provided with an interface where they sequentially sampled the outcomes predicted by an underlying model. In the first, we tested responses to seven well-known probabilistic puzzles. The second involved simulating the predictions of an estimated regression model, given one's choices, in order to make investment decisions. The results of both experiments are unequivocal. Experience obtained through simulations led to far more accurate inferences than attempts at analysis. Also, the participants preferred using the experiential methodology over analysis. Moreover, when aided by simulation, participants who were naïve with respect to probabilistic reasoning performed as well as those with university training in statistical inference. The results support our suggestion that the authors of empirical papers supplement the outputs of their analyses with simulation models that allow decision makers to "go beyond the information given" and "experience" the outcomes of the model given their inputs.

Although our suggestion would impose an additional burden on authors, it would reduce both effort and misinterpretation on the part of readers, and would make any empirical article a more accessible scientific product. Moreover, it has the potential to correct other statistical misinterpretations that were not identified by our study. As such, we believe that our suggestion goes a long way to toward increasing our understanding of economic phenomena. At the same time, it also calls for additional research into understanding when and why different presentation formats lead to misinterpretation.

5 For example, by following the link http://www.econ.upf.edu/~soyer/Emre_Soyer/Econometrics_Project.html, the reader can investigate many questions concerning the two regression set-ups that we examined in this paper, and can also experience simulated outcomes.

In addition to suggesting changes in the way in which statistical results are reported in journals in order to produce better inferences, our results also have implications for the teaching of statistical techniques. First, textbooks should provide a better coverage of the way to report statistical results, as well as instructions as to how to make probabilistic predictions. Even a cursory examination of leading textbooks shows that the topic of reporting currently receives little attention, while decision making is only considered through the construction of confidence intervals around predicted outcomes.

Together with estimating average effects, evaluating the predictive ability of economic models should become an important component of econometrics teaching. Indeed, if linked to the development and use of simulation methods, it could become a most attractive (and illuminating) part of any econometrics syllabus.

Finally, we note that scientific knowledge advances to the extent that we are able to forecast and control different phenomena. However, if we cannot make appropriate probabilistic statements about our predictions, our ability to assess our level of knowledge accurately is seriously compromised.

Acknowledgments

The authors are particularly grateful to the economists who took the time to answer the survey. In addition, they are indebted to several colleagues for excellent advice on matters ranging from planning the study, to its implementation, and the paper itself. These include Manel Baucells, Michael Greenacre, Gaël Le Mens, Stephan Litschig, Johannes Mueller-Trede, Omiros Papaspiliopoulos, Gueorgui Kolev, and especially Nick Longford, Rosemarie Nagel, and Thijs van Rens. We also thank the late Arnold Zellner for his comments on the work, and dedicate the paper in honor of his memory. Comments made at seminars at Universitat Pompeu Fabra, at the Spanish Economic Association Symposium (SAEe) 2010, and at the Royal Economic Society Ph.D. Meeting 2010 were also much appreciated. The research was supported by grants SEJ2006-14098 and EC02009-09834 from the Spanish Ministerio de Ciencia e Innovación.

Appendix A. Rationale for answers to the four questions

A.1. Preliminary comments

We test whether or not decision makers who are knowledgeable about regression analysis evaluate the unpredictability of an outcome correctly, given the standard presentation of linear regression results in an empirical study. To isolate the effects of a possible misperception, we created a basic specification. In this hypothetical situation, a continuous variable X causes an outcome Y, and the effect of one more X is estimated to be almost exactly equal to 1. The majority of the fluctuation in Y is due to a random disturbance uncorrelated with X, which is normally and independently distributed, with constant variance. Hence, the decision maker knows that all of the assumptions of the classical linear regression model hold (see, e.g., Greene, 2003).

A.2. Answers to questions 1 and 2

In the first two questions, participants are asked to advise a hypothetical individual who desires to have a certain level of control over the outcomes. This corresponds to the desire to obtain a certain amount of Y through some action X. The first question reflects the desire to obtain a positive outcome, whereas the second reflects the desire to be better off with respect to an alternative of no-action. If one considers only averages, the estimation results suggest that an individual should expect the relationship between X and Y to be one to one. However, when could an individual claim that a certain outcome has occurred because of their actions, and is not merely due to chance? How much does chance have to say in the realization of an outcome? The answers to these questions depend on the standard deviation of the estimated residuals (SER).

In a linear regression analysis, SER^2 corresponds to the variance of the dependent variable that cannot be explained by the independent variables, and is captured by the statistic $(1-R^2)$. In Conditions 1 and 3, this is given as 50%. One can compute the SER using the $(1-R^2)$ statistic and the variance of Y:

$$\begin{aligned}\mathrm{SDER} &= se(\hat{e}) = \sqrt{\mathrm{Var}(Y)(1-R^2)}\\ &= \sqrt{(40.78^2)(0.5)} \approx 29.\end{aligned} \tag{A.1}$$

The answer to the first question can be calculated approximately by constructing a one-sided 95% confidence interval using Eq. (A.1). We are looking for a value of X where

$$\begin{aligned}&\mathrm{Prob}\left(Z > -\frac{\hat{C}+\hat{\beta}X}{se(\hat{e})}\right)\\ &\quad= \mathrm{Prob}\left(Z > -\frac{0.32+1.001X}{29}\right)\\ &\quad= 0.95, \quad \text{where } Z \sim N(0,1).\end{aligned} \tag{A.2}$$

Thus, to obtain a positive payoff with 95% probability, an individual has to choose:

$$X = \frac{(1.645 * 29 - 0.32)}{1.001} \approx 47. \tag{A.3}$$

The answer to the second question requires one additional calculation. Specifically, we need to know the standard deviation of the difference between two random variables, that is

$$(Y_i \mid X_i = x_i) - (Y_i \mid X_i = 0), \quad \text{where } x_i > 0. \tag{A.4}$$

We know that $(Y_i \mid X_i)$ is an identically, independently and normally distributed random error with an estimated standard deviation of 29. Given that different and independent shocks occur for different individuals and actions, the standard deviation of Eq. (A.4) becomes:

$$\begin{aligned}&\sqrt{\mathrm{Var}[(Y_i \mid X_i = x_i) - (Y_i \mid X_i = 0)]}\\ &\quad= \sqrt{\mathrm{Var}(Y_i \mid X_i = x_i) + \mathrm{Var}(Y_i \mid X_i = 0)}\\ &\quad= \sqrt{29^2 + 29^2} \approx 41.\end{aligned} \tag{A.5}$$

Thus, the answer to question 2 is:

$$X = \frac{(1.645 * 41 - 0.32)}{1.001} \approx 67. \quad \text{(A.6)}$$

Similar reasoning is involved for Condition 2 (and thus also Conditions 4 and 6). For these conditions, the equivalent of Eq. (A.1) is

$$\begin{aligned} \mathrm{SDER} = se(\hat{e}) &= \sqrt{\mathrm{Var}(Y)(1 - R^2)} \\ &= \sqrt{(59.25^2)(0.75)} \approx 51, \end{aligned} \quad \text{(A.7)}$$

such that the answer to question 1 is:

$$X = \frac{(1.645 * 51 - 0.62)}{1.02} \approx 82. \quad \text{(A.8)}$$

As for question 2, we need to find out about Eq. (A.4) in this condition:

$$\begin{aligned} &\sqrt{\mathrm{Var}(Y_i \mid X_i = x_i) + \mathrm{Var}(Y_i \mid X_i = 0)} \\ &\quad = \sqrt{51^2 + 51^2} \approx 72, \end{aligned} \quad \text{(A.9)}$$

so that the answer to question 2 in Condition 2 becomes:

$$X = \frac{(1.645 * 72 - 0.62)}{1.02} \approx 116. \quad \text{(A.10)}$$

A.3. Answers to questions 3 and 4

Here, we inquire about the way in which decision makers weight the different sources of uncertainty within the dependent variable. The answers to these questions provide insights as to whether or not the typical presentation of the results leads the participants to consider that the fluctuation around the estimated coefficient is a larger source of uncertainty in the realization of Y than it really is.

Question 3 asks about the probability of obtaining an outcome above the lower-bound of the 95% confidence interval of the estimated coefficient, given a value of $X = 1$.

In Conditions 1, 3 and 5, the lower-bound is 0.936. We can find an approximate answer to this question using the estimated model and the SER from Eq. (A.1), that is

$$\begin{aligned} &\Pr(Y_i > 0.936 \mid X_i = 1) \\ &\quad = \Pr(\hat{C} + \hat{\beta}X_i + \hat{e} > 0.936 \mid X_i = 1) \\ &\quad = \Pr(\hat{e} > 0.936 - \hat{C} - \hat{\beta}X_i \mid X_i = 1) \\ &\quad = \Pr\left(\frac{\hat{e}}{se(\hat{e})} > \frac{0.936 - \hat{C} - \hat{\beta}X_i}{se(\hat{e})} \mid X_i = 1\right) \\ &\quad = 1 - \Phi\left(\frac{0.936 - 0.32 - 1.001}{29}\right) \\ &\quad = 1 - \Phi(-0.013) \approx 0.51, \end{aligned} \quad \text{(A.11)}$$

where Φ is the cumulative standard normal distribution.

Question 4 asks about the probability of obtaining an outcome above the point estimate, given a value of $X = 1$. In Conditions 1, 3 and 5, the point estimate is 1.001. We can use similar calculations in order to obtain an answer.

$$\begin{aligned} &\Pr(Y_i > 1.001 \mid X_i = 1) \\ &\quad = \Pr(\hat{C} + \hat{\beta}X_i + \hat{e} > 1.001 \mid X_i = 1) \\ &\quad = \Pr(\hat{e} > 1.001 - \hat{C} - \hat{\beta}X_i \mid X_i = 1) \\ &\quad = \Pr\left(\frac{\hat{e}}{se(\hat{e})} > \frac{1.001 - \hat{C} - \hat{\beta}X_i}{se(\hat{e})} \mid X_i = 1\right) \\ &\quad = 1 - \Phi\left(\frac{1.001 - 0.32 - 1.001}{29}\right) \\ &\quad = 1 - \Phi(-0.01) \approx 0.5. \end{aligned} \quad \text{(A.12)}$$

For questions 3 and 4 of Condition 2 (and thus also 4 and 6), we follow a similar line of reasoning, using the appropriate estimates. Thus, for question 3,

$$\begin{aligned} &\Pr(Y_i > 0.911 \mid X_i = 1) \\ &\quad = \Pr(\hat{C} + \hat{\beta}X_i + \hat{e} > 0.911 \mid X_i = 1) \\ &\quad = \Pr(\hat{e} > 0.911 - \hat{C} - \hat{\beta}X_i \mid X_i = 1) \\ &\quad = \Pr\left(\frac{\hat{e}}{se(\hat{e})} > \frac{0.911 - \hat{C} - \hat{\beta}X_i}{se(\hat{e})} \mid X_i = 1\right) \\ &\quad = 1 - \Phi\left(\frac{0.911 - 0.61 - 1.02}{51}\right) \\ &\quad = 1 - \Phi(-0.015) \approx 0.51, \end{aligned} \quad \text{(A.13)}$$

and for question 4,

$$\begin{aligned} &\Pr(Y_i > 1.02 \mid X_i = 1) \\ &\quad = \Pr(\hat{C} + \hat{\beta}X_i + \hat{e} > 1.02 \mid X_i = 1) \\ &\quad = \Pr(\hat{e} > 1.02 - \hat{C} - \hat{\beta}X_i \mid X_i = 1) \\ &\quad = \Pr\left(\frac{\hat{e}}{se(\hat{e})} > \frac{1.02 - \hat{C} - \hat{\beta}X_i}{se(\hat{e})} \mid X_i = 1\right) \\ &\quad = 1 - \Phi\left(\frac{1.02 - 0.61 - 1.02}{51}\right) \\ &\quad = 1 - \Phi(-0.01) \approx 0.5. \end{aligned} \quad \text{(A.14)}$$

Appendix B. Histograms for the answers to Conditions 2, 3, 4 and 6

See Figs. B.1–B.4.

Appendix C. Detailed experimental data for Conditions 1–4

See Table C.1.

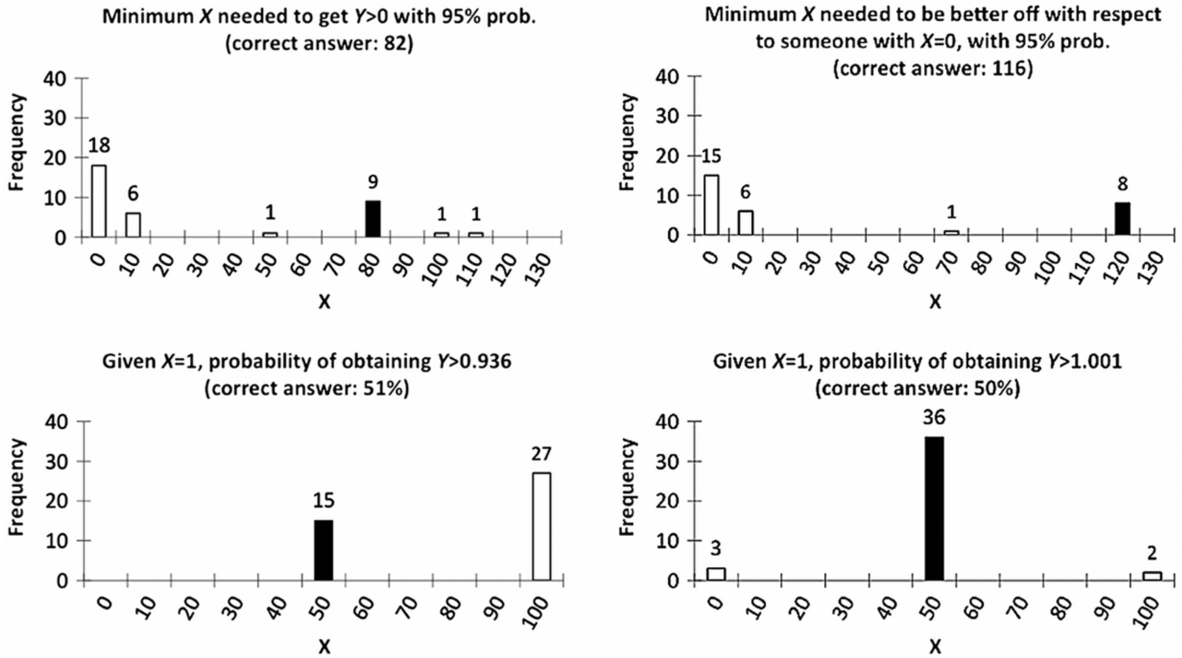

Fig. B.1. Histograms for the responses to Condition 2. The top-left figure shows answers to question 1, the one on the top-right shows answers to question 2, the one on the bottom-left those to question 3, and the one on the bottom-right those to question 4. Each histogram also displays the question and the approximate correct answer. The dark column identifies the responses that we considered correct. Above each column is the number of participants who gave that particular answer. There were 36, 30, 42 and 41 responses to questions 1–4, respectively.

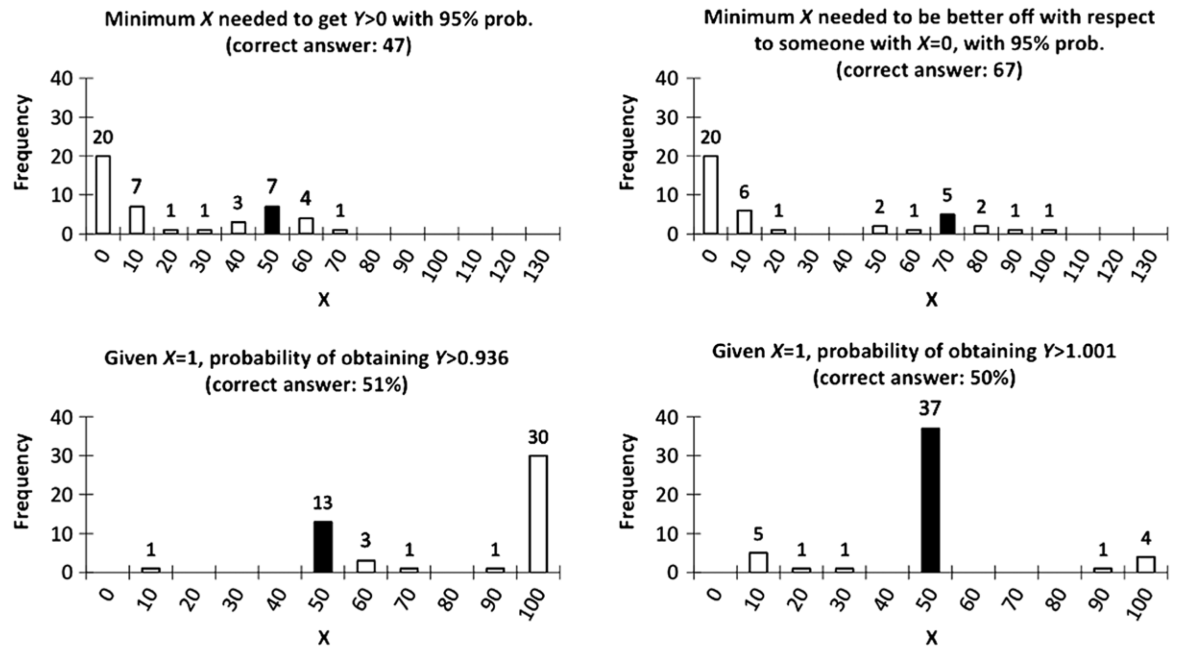

Fig. B.2. Histograms for the responses to Condition 3. The top-left figure shows answers to question 1, the one on the top-right shows answers to question 2, the one on the bottom-left those to question 3, and the one on the bottom-right those to question 4. Each histogram also displays the question and the approximate correct answer. The dark column identifies the responses that we considered correct. Above each column is the number of participants who gave that particular answer. There were 44, 39, 49 and 49 responses to questions 1–4, respectively.

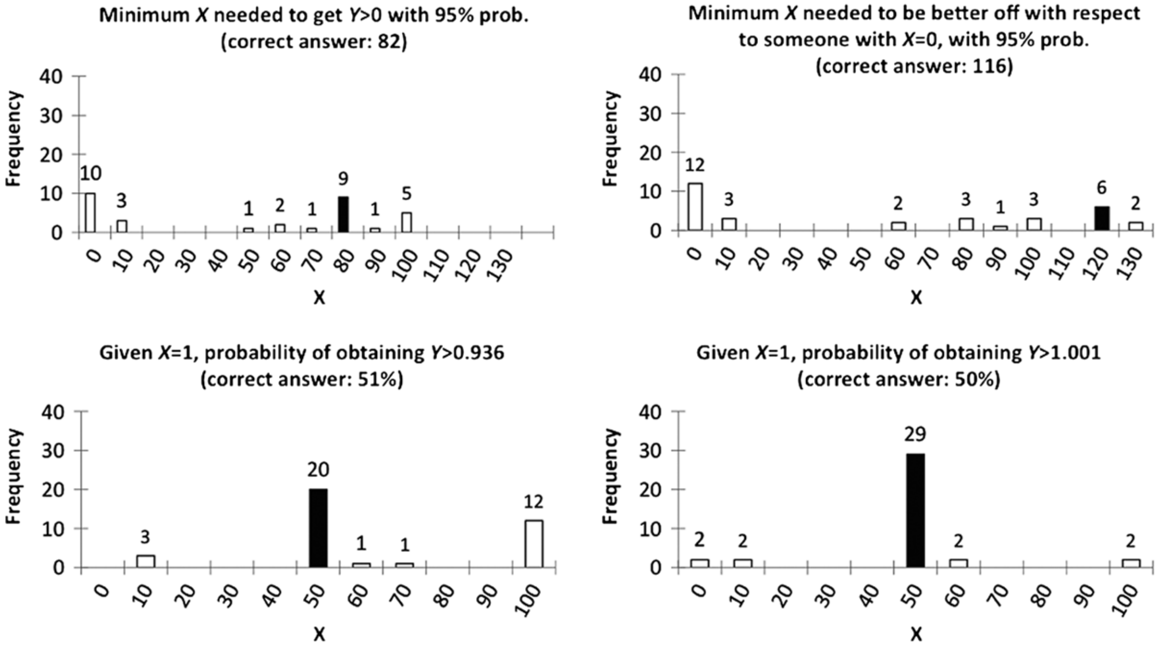

Fig. B.3. Histograms for the responses to Condition 4. The top-left figure shows answers to question 1, the one on the top-right shows answers to question 2, the one on the bottom-left those to question 3 and the one on the bottom-right those to question 4. Each histogram also displays the question and the approximate correct answer. The dark column identifies the responses that we considered correct. Above each column is the number of participants who gave that particular answer. There were 32, 32, 37 and 37 responses to questions 1–4, respectively.

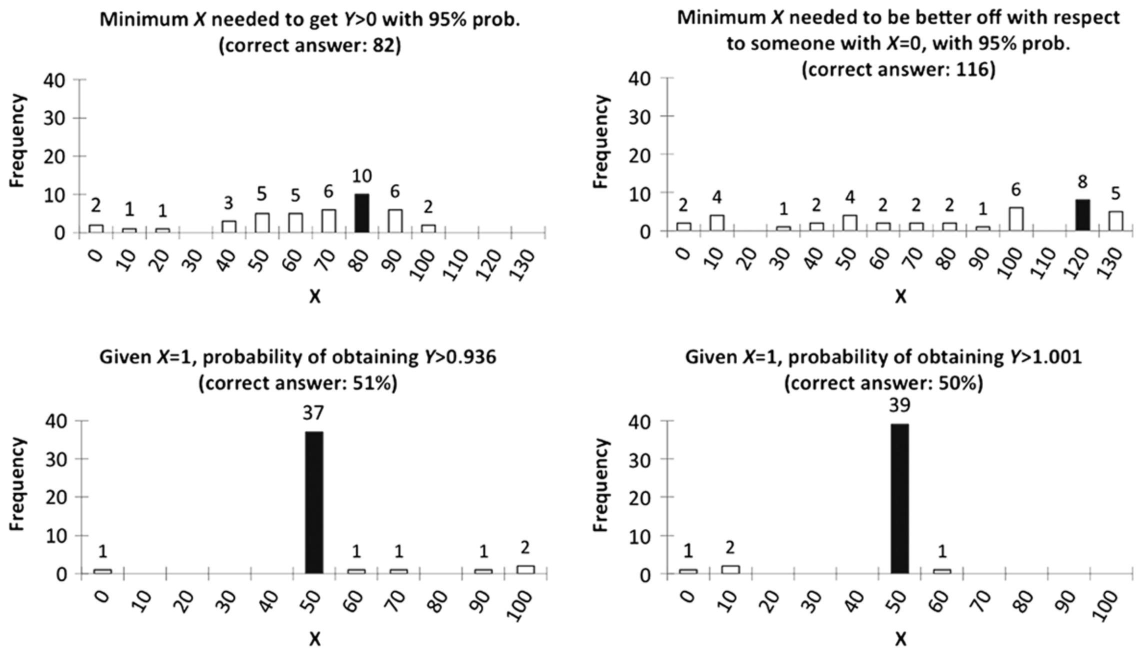

Fig. B.4. Histograms for the responses to Condition 6. The top-left figure shows answers to question 1, the one on the top-right to shows answers question 2, the one on the bottom-left those to question 3 and the one on the bottom-right those to question 4. Each histogram also displays the question and the approximate correct answer. The dark column identifies the responses that we considered correct. Above each column is the number of participants who gave that particular answer. There were 41, 39, 43 and 43 responses to questions 1–4, respectively.

Table C.1
Relationships between training, experience and responses in Conditions 1–4 (the number of respondents with correct answers is given in parentheses).

Condition	1	2	3	4	Total over four conditions	Percentage of respondents with correct answers
Position						
Professor	17 (4)	14 (5)	19 (6)	18 (11)	68 (26)	38
Associate Professor	8 (2)	7 (3)	12 (4)	10 (8)	37 (17)	46
Assistant Professor	12 (5)	18 (4)	16 (6)	9 (2)	55 (17)	31
Senior Lecturer	0 (0)	2 (1)	1 (0)	0 (0)	3 (1)	33
Lecturer	6 (1)	4 (0)	1 (0)	0 (0)	12 (1)	8
Post-Doctoral Researcher	2 (0)	0 (0)	0 (0)	0 (0)	2 (0)	0
Total	45 (12)	45 (13)	49 (13)	38 (21)	177 (62)	35
Research fields						
Econometrics	14 (6)	11 (6)	10 (5)	14 (8)	49 (25)	51
Labor economics	12 (5)	11 (2)	14 (3)	10 (7)	47 (17)	36
Monetary economics	5 (1)	2 (0)	5 (2)	2 (0)	14 (3)	21
Financial economics	4 (1)	5 (3)	4 (3)	3 (2)	16 (9)	56
Behavioral economics	3 (1)	7 (2)	2 (1)	3 (0)	15 (4)	27
Developmental economics	8 (1)	2 (1)	9 (3)	5 (1)	24 (6)	25
Health economics	4 (0)	3 (0)	5 (1)	1 (1)	13 (2)	15
Political economy	3 (1)	5 (1)	7 (3)	4 (2)	19 (7)	37
Public economics	9 (1)	6 (1)	10 (4)	8 (6)	33 (12)	36
Environmental economics	1 (0)	2 (1)	3 (0)	2 (1)	8 (2)	25
Industrial organization	2 (1)	6 (1)	6 (1)	2 (1)	16 (3)	19
Game theory	4 (1)	1 (1)	4 (1)	5 (2)	14 (5)	36
International economics	6 (2)	6 (0)	7 (1)	2 (1)	21 (4)	19
Macroeconomics	9 (2)	9 (2)	13 (2)	6 (5)	37 (11)	30
Microeconomics	11 (2)	4 (2)	11 (5)	7 (4)	33 (13)	39
Economic history	2 (0)	2 (0)	6 (3)	2 (1)	12 (4)	33
Statistics	3 (1)	4 (4)	1 (1)	1 (1)	11 (7)	64
Other	0 (0)	0 (0)	1 (1)	0 (0)	1 (1)	100
Use of regression analysis						
Never	7 (1)	5 (0)	11 (7)	11 (5)	34 (13)	38
Some	11 (4)	16 (6)	17 (0)	10 (5)	54 (15)	28
Often	16 (4)	14 (5)	7 (2)	7 (6)	44 (17)	39
Always	5 (3)	5 (1)	8 (4)	6 (2)	24 (10)	42
Total	39 (12)	40 (12)	43 (13)	34 (18)	156 (55)	35
Average minutes spent	12 (10.9)	10.6 (12.6)	7.4 (11.2)	7.5 (7.4)	8.1 (10.2)	8.1
Std. dev.	12 (9.4)	7.8 (9)	7.1 (12.3)	5.3 (5.2)	7.7 (9)	7.7

References

Armstrong, J. S. (2007). Significance tests harm progress in forecasting. *International Journal of Forecasting*, *23*, 321–327.

Baltagi, B. H. (2007). Worldwide econometrics rankings: 1989–2005. *Econometric Theory*, *23*(5), 952–1012.

Camerer, C. F. (2000). Prospect theory in the wild: evidence from the field. In D. Kahneman, & A. Tversky (Eds.), *Choice, values, and frames* (pp. 288–300). New York, NY: Russell Sage Foundation & Cambridge University Press.

Camerer, C. F., & Hogarth, R. M. (1999). The effects of financial incentives in experiments: a review and capital-labor-production framework. *Journal of Risk and Uncertainty*, *19*, 7–42.

Carhart, M. (1997). On persistence in mutual fund performance. *Journal of Finance*, *52*(1), 57–82.

Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M., & Woloshin, S. (2007). Helping doctors and patients make sense of health statistics. *Psychological Science in the Public Interest*, *8*(2), 53–96.

Goldstein, D. G., & Taleb, N. N. (2007). We don't quite know what we are talking about when we talk about volatility. *Journal of Portfolio Management*, *33*(4), 84–86.

Greene, W. H. (2003). *Econometric analysis* (5th ed.). Upper Saddle River, NJ: Prentice Hall.

Gujarati, D. N., & Porter, D. (2009). *Basic econometrics*. New York: McGraw-Hill Irwin.

Hendry, D. F., & Nielsen, B. (2007). *Econometric modeling: a likelihood approach*. NJ: Princeton University Press.

Hogarth, R. M., & Soyer, E. (2011). Sequentially simulated outcomes: kind experience vs. non-transparent description. *Journal of Experimental Psychology: General*, *140*, 434–463.

Hubbard, R., & Armstrong, J. S. (1994). Replications and extensions in marketing—rarely published but quite contrary. *International Journal of Research in Marketing*, *11*, 233–248.

Jensen, M. C. (1968). The performance of mutual funds in the period 1945–1964. *Journal of Finance*, *23*(2), 389–416.

Judge, G. G., Griffiths, W., Hill, C. R., & Lee, T. C. (1985). *Theory and practice in econometrics*. New York: Wiley.

Kahneman, D., & Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica*, *47*, 263–291.

Lawrence, M., & Makridakis, S. (1989). Factors affecting judgmental forecasts and confidence intervals. *Organizational Behavior and Human Decision Processes*, *42*, 172–187.

McCloskey, D. N., & Ziliak, S. T. (1996). The standard error of regressions. *Journal of Economic Literature*, *34*, 97–114.

Schwab, A., & Starbuck, W. H. (2009). Null-hypothesis significance testing in behavioral and management research: we can do better. In D. Bergh, & D. Ketchen (Eds.), *Research methodology in strategy and management*: *Vol. 5* (pp. 29–54). Oxford, UK: Elsevier.

Simon, H. A. (1978). Rationality as process and product of thought. *American Economic Review*, *68*(2), 1–16.

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: improving decisions about health, wealth, and happiness*. New Haven, CT: Yale University Press.
Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin*, *76*, 105–110.
Wooldridge, J. M. (2008). *Introductory econometrics: a modern approach* (3rd ed.). International Student Edition, Thomson, South Western.
Zellner, A. (1984). Posterior odds ratios for regression hypotheses: general considerations and some specific results. In A. Zellner (Ed.), *Basic issues in econometrics* (pp. 275–305). Chicago, IL: The University of Chicago Press.
Zellner, A. (2004). To test or not to test and if so, how? Comments on "size matters". *Journal of Socio-Economics*, *33*, 581–586.
Ziliak, S. T., & McCloskey, D. N. (2004). Size matters: the standard error of regressions in the American Economic Review. *Journal of Socio-Economics*, *33*, 527–546.
Ziliak, S. T., & McCloskey, D. N. (2008). *The cult of statistical significance: how the standard error costs us jobs, justice, and lives*. Ann Arbor: University of Michigan Press.

Emre Soyer is a Ph.D. student in the Department of Economics & Business at Universitat Pompeu Fabra, Barcelona. A graduate of Koc University (Istanbul, Turkey) and the University of Nottingham (U.K), he is interested in ways of structuring situations so as to help unleash human potential across a wide range of areas, ranging from simple decision problems to the content of educational programs.

Robin M. Hogarth is an ICREA Research Professor in the Department of Economics & Business at Universitat Pompeu Fabra, Barcelona. He has previously held appointments at INSEAD, London Business School, and the University of Chicago. He has published several books (most recently *Dance with Chance* with Spyros Makridakis and Anil Gaba) and many articles in psychology, management, and economics on topics related to human decision making. He is a past President of both the Society for Judgment and Decision Making and the European Association for Decision Making.