Econometrics and decision making: Effects of presentation mode *

Emre Soyer[1] & Robin M. Hogarth[1,2]

[1]Universitat Pompeu Fabra, Department of Economics & Business, Barcelona

[2]ICREA, Barcelona

emre.soyer@upf.edu

robin.hogarth@upf.edu

Abstract

Much of empirical economics involves regression analysis. However, does the presentation of results affect economists' ability to make inferences for decision making purposes? In a survey, 257 academic economists were asked to make probabilistic inferences on the basis of the outputs of a regression analysis presented in a standard format. Questions concerned the distribution of the dependent variable conditional on known values of the independent variable. However, many respondents underestimated uncertainty by failing to take into account the standard deviation of the estimated residuals. The addition of graphs did not substantially improve inferences. On the other hand, when *only* graphs were provided (i.e., with no statistics), respondents were substantially more accurate. We discuss implications for improving practice in reporting results of regression analyses.

Keywords: regression analysis; presentation formats; probabilistic predictions; graphs
JEL: C01; C20; C53; Y10

Version: February 11, 2010

* 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, its implementation, and the paper itself. These include Rosemarie Nagel, Michael Greenacre, Gaël Le Mens, Johannes Mueller-Trede, Omiros Papaspiliopoulos, Gueorgui Kolev, and especially Nick Longford. The research was also supported by grant SEJ2006-14098 from the Spanish Ministerio de Ciencia y Innovación.

I. Introduction

The leading journals in economics are the life-blood of the profession in that they report high quality, peer-reviewed research on the latest theoretical and empirical developments. As such, the findings are of great interest to a wide range of social scientists ranging from theoreticians to practically-minded economists working on applied problems. In dealing with almost any issue, these scientists all face two questions: (1) Which variables are important in explaining economic outcomes? and (2) How important are the variables that have been identified?

Our contention in this paper is that the manner in which the results of empirical analyses are presented in leading economics journals hinders the ability of economists to answer these questions. This is especially the case when the work requires interpretation from a decision making perspective as required, for example, in policy analysis.

Whereas it can be argued that *how* information is presented should not affect rational interpretation and analysis, there is abundant psychological evidence demonstrating presentation effects. Many studies have shown, for example, how subtle changes in questions designed to elicit preferences are subject to so-called framing and other contextual influences (see, e.g., Kahneman & Tversky, 1979; Hogarth, 1982; Tversky & Kahneman, 1986). Moreover, these have been reported in both controlled laboratory conditions and field studies involving appropriately motivated experts (Camerer, 2000; McNeil et al., 1982; Thaler & Sunstein, 2008). 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 et al., 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 miscommunication of statistical information are common in everyday situations, resulting in misperceptions and irrational decisions. Their work focuses mainly on the fields of medicine and law, where in particular situations, doctors, lawyers and judges fail to communicate crucial statistical information appropriately thereby leading to biased judgments that impact negatively on others.

We examine how economists communicate statistical information among themselves. Specifically, we note that much work in empirical economics involves the technique of regression analysis. However, when we asked a large sample of economists to use the standard reported outputs of regression analysis to make probabilistic inferences for decision making purposes, they experienced considerable difficulty. The reason, we believe, is that current reporting practices focus attention on the uncertainty surrounding 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 is directed appropriately – by, for example, graphical as opposed to tabular means – the quality of our respondents' inferences increases dramatically.

This paper is organized as follows. In the next section (II), we provide some background on the practice and evolution of reporting empirical results in journals in economics and contrast this with developments in other social sciences. Subsequently (section III) we provide information concerning the survey we conducted with economists that involved answering four decision-oriented questions based on a standard format for reporting results of regression analysis. There were six different conditions designed to assess differential effects due to model fit (R^2) and different forms of graphical presentation (with and without accompanying statistics). In section IV, we present our results. In brief, these show that the typical presentation format – that highlights regression coefficients and their standard errors – leads respondents to ignore the level of predictive uncertainty in the model that is captured by the standard deviation of the estimated residuals. As a consequence, uncertainty is grossly underrated. Moreover, adding graphs does little to ameliorate the situation. On the other hand, the provision of graphs *alone* – without regression statistics – does lead to more accurate inferences. The implications of our findings are discussed in the concluding section, V.

II. Current practice

There are many sources of empirical analyses and findings in economics. To obtain a representative sample of current "best" practice, we selected 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). Many articles published in these journals are empirical. Over 70% of the empirical analyses use variations of regression analysis of which 75% have linear specifications. This suggests that regression analysis is the most prominent tool used by economists to test hypotheses and identify relations among economic and social variables. The use of regression analysis is also common in other fields such as political science, psychology, and education (Gall et al., 1996; Willson, 1980; Pedhazur, 1997).

Empirical studies published in economics journals follow a common procedure to display and evaluate results. In a typical study, authors provide a table that displays the descriptive statistics of the data sample used in the analysis. Before or after the display of this information, they describe the specification of the model on which the analysis is based. Then the regression results are provided 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 R^2. Table 1 summarizes these details with respect to the sample of studies referred to above. This shows that, apart from the regression coefficients and their standard errors (or *t*-statistics), there is not much agreement on what else should be reported. The data suggest, therefore, that – as a group – economists probably understand well the inferences that can be made about regression coefficients or the *average* impact of manipulating an independent variable; however, their ability to make inferences about other probabilistic implications is possibly less well developed (e.g., predicting individual outcomes conditional on specific inputs).

Insert Table 1 about here

It is not clear when, how, and why the above manner of presenting regression results in publications emerged. No procedure is made explicit in the submission guidelines for the highly ranked journals. Moreover, popular econometric textbooks, such as Greene (2003), Judge (1985) and Gujarati (1988) do not explain specifically how to present estimation results or how to use them for decision making. An exception is Wooldridge (2008), who dedicates several sections to issues of presentation. His outline suggests that a good summary of an analysis consists of a table with selected coefficient estimates and their standard errors, R^2 statistic, constant, and the number of observations. Indeed, this seems to be consistent with today's practice, as more than 60% of the articles in Table 1 follow a similar procedure.

Publications from the 1950's and 1960's display and discuss results in a fashion similar to today, even though in that period empirical analyses constituted a minority of the studies published in QJE, AER, JPE and RES (about 25% vs. today's 50%). Hence, despite the considerable growth and advances in empirical work in economics – due undoubtedly to developments in computational technology – the content of the display and discussion of results has remained remarkably constant over time. Widely used statistical software, such as STATA, RATS, SPSS or MS-EXCEL display statistics and regression results in a way analogous to the tables used in published papers, but provide a larger number of statistics, out of which only a handful are selected by the authors and discussed in the publications.

The presentation of statistical results has been debated over the years in different fields of research and has led to several innovations. For example, augmenting significance tests with effect size became a common practice in differential psychology in the 1980's. This has also been adopted by empirical scientists in fields such as sociology and political science where it is used to analyze treatment effects in experiments or to conduct meta-analyses. *Psychological Science*, the flagship journal of the research oriented Association for Psychological Science is probably the leader in advocating specific statistical reporting practices in the social sciences. For example, its "Information for Contributors" 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 economics, McCloskey and Ziliak (1996) provided an illuminating study of statistical practice based on articles published in AER in the 1980s. They demonstrated widespread confusion in the interpretation of statistical results due to confounding the concept of statistical significance with notions of economic or substantive significance. Too many "results" were dependent on whether *t* or other statistics simply exceeded arbitrarily defined limits. Although we have not conducted any formal analysis, we have little reason to believe that the situation has changed since the 1980s.

Practice in empirical finance provides an interesting exception. In this field, once statistical analysis has identified a variable as "important" in affecting, say, stock returns, it is standard to assess "how important" by evaluating the performance of simulated stock portfolios that use the variable. In short, by using the information contained in the variable when it becomes available to market participants, what increase in future portfolio value can be gained? For applications of this method to evaluating performance of active fund managers, see Jensen (1968) and Carhart (1997).

At the beginning of the 1990's, several researchers in political science showed interest in the predictive power of published models and discussed the effectiveness of the statistics reported in providing individuals with reliable information for decision making. Specifically, King (1990a, 1990b), Krueger and Lewis-Beck (2007), Lewis-Beck and Skalaban (1990a, 1990b, 1991) narrowed down the methodological aspects of this issue to a rivalry between R^2 and the Standard Deviation of the Estimated Residuals (SDER).[1]

The R^2 advocates argue that, as a bounded and standardized quantity, this statistic is the best option to describe the fit of a model and a useful measure of the relative predictive abilities of different specifications. The SDER advocates, on the other hand, argue that R^2 statistics may vary considerably among different samples whereas SDER provides information on the degree of predictability in the metric of the dependent variable. Despite the different points of view, there is agreement that both statistics aid

[1] We use the initials SDER to indicate the standard deviation of the estimated residuals. Some sources refer to this as the Standard Error of Estimate or SEE (see RATS), some others as root Mean Squared Error or root-MSE (see STATA). Wooldridge (2008) calls it the Standard Error of Regression (SER) defining it as "an estimator of the standard deviation of the error term."

decision makers in understanding different aspects of the level of predictability associated with the analyses.

Table 1 shows that SDER is practically absent from the presentation of results. Less than 10% of the studies provide it. R^2 seems to be the prevalent statistic reported to provide an idea of model fit. This is the case for 80% of the published articles with a linear specification.

It is difficult to assess the magnitude of the uncertainty in a regression analysis without knowing both the R^2 statistic and the standard deviation of the dependent variable. However, Table 1 shows that more than 40% of the publications in our sample that utilize a linear regression analysis do not provide information on at least one of these statistics. Hence, a decision maker who is consulting these studies cannot infer much about the cloud of data points on which the regression line is fit. Alternatively, a scatter plot would be essential to perceive the degree of uncertainty. However, less than 40% of publications in our sample of regression analyses provide a graph with actual observations.

Given the prevalence of empirical analyses and their potential use for decision making, debates about the appropriate way to present results are important. However, none of these previous debates is based on systematic evidence of how knowledgeable individuals use the current tools for making probabilistic inferences, and how – given an estimated model – specific statistics and different ways of presenting results affect judgment. The purpose of this investigation is to illuminate this issue.

III. The survey

Goal and design

The goal of our survey was to investigate how knowledgeable individuals (economists) interpret specific decision making implications of the standard output of a regression analysis. We applied the following criteria to select the survey questions. First, we provided information about a well-specified model that met the underlying assumptions of regression analysis. Second, the model was straightforward in that it had only one independent variable. Third, all the information necessary to solve the problems we posed was available from the output provided in the analysis. Fourth, although sufficient information was available, respondents needed to apply knowledge about statistical inference to make the calculations necessary to answer the questions. That is, respondents had to go beyond just the information provided.

This last criterion is the most demanding because, we believed, 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). However, these statements are essential for decision making. For example, imagine that a regression has been carried out showing the relation in a specific population between annual earnings (dependent variable) and years-spent-at-school (independent variable). Now, consider a specific person from the same population who has spent k years-at-school and is considering spending an additional year (i.e., by increasing k to $k+1$). From a decision perspective, obvious questions center on the probable effects of this additional year on earnings. What, for example, is the probability that an extra year at school will lead to earnings in excess of a specific level? This is precisely the kind of question we ask our respondents.

The design of our study required that respondents answer four such questions after being provided with information about an underlying regression analysis. Our survey involved six conditions and Figures 1 and 2 report the information provided to the respondents for Conditions 1 and 2, respectively. We make three comments about these set-ups. First, the information provided is similar in form and content to the outputs of many regression analyses reported in the economic literature (and consistent with the prescriptions of Wooldridge, 2008). Second, all the assumptions of regression are satisfied in a way that might not be strictly possible with empirical data (thus the estimated model contains information concerning *all* uncertainties involved in prediction). Third, the main difference between Conditions 1 and 2 lies in the overall "fit" of the regression model. In Condition 1, R^2 is 0.50; in Condition 2, it is 0.25.

Insert Figures 1 and 2 about here

A possible critique of Conditions 1 and 2 is that some economists would also like information in the form of the bivariate scatter-plot of the dependent and independent variables as well the standard deviation of the estimated residuals (indeed, as noted above, in some reports both can be found). Conditions 3 and 4 were the same as Conditions 1 and 2, respectively, except that this additional information was included – see Figures 3 and 4.

Insert Figures 3 and 4 about here

Finally, we explored what would happen if, instead of the usual reporting of regression statistics, respondents were forced to respond to our questions by simply consulting graphs. Thus, 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 Figures 3 and 4.[2]

In summary, differences between each of Conditions 1 and 2, 3 and 4, and 5 and 6, all reflect differences in variance explained (R^2 of 0.50 versus R^2 of 0.25). Conditions 3 and 4 add graphs and SDER to the information in Conditions 1 and 2. And the results of the regression analyses are limited to graphs in Conditions 5 and 6.

It is important to add that in published papers, results are also discussed verbally. These detailed discussions, which are mostly confined to certain coefficient estimates and their significance, might distract decision makers from the uncertainties about outcomes. None of our conditions involve such discussions. Furthermore, in publications, some discussions describe relations among variables using attributes, such as “strong”, “weak”, “determinant”, “predictor” etc. relying solely on the statistical significance of coefficient estimates. It might be possible that these explanations, which are a part of the current rhetoric used in reporting results, frame decision makers into believing that the results imply even more predictive power than is the case.

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

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. 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>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 in questions 3 and 4 is (0.911, 1.130), and we ask respectively about the probabilities of obtaining Y>0.911 and Y>1.02, given X=1. All four questions are reasonable in that they seek answers to questions that would be of interest to decision makers. However – and as noted above – they are not the types of questions that reports in economic journals usually lead readers to pose. They therefore test a respondent's ability to reason correctly in a statistical manner given the information provided. In Appendix A, we provide the rationale behind the correct answers.

Respondents and method

We sent web-based surveys to faculty members in economics departments worldwide. One hundred and thirteen departments were randomly selected from a list of 150 compiled by Baltagi (2007, Table 3) who ranks economics departments of universities worldwide by their econometric publications between 1989 and 2005.[3] Within each department, we randomly selected up to 32 faculty members. We ordered

[3] We stopped sampling universities once we had at least 30 individual responses for all questions asked. A few universities were not included in our sample because their web pages did not facilitate accessing potential respondents. This was more frequent for non-US universities. For reasons of confidentiality, we do not identify any of these universities.

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] We did not offer respondents any economic incentives for participation but note that it is not clear what difference such incentives would make in the present case (Camerer & Hogarth, 1999).

When starting our investigation we had little idea as to how many economists would actually respond to our survey. We therefore started collecting data for Conditions 1 and 2 from a group of economics departments that are not in our final sample to estimate how many requests would be needed to achieve sample sizes of between 30 and 40 in each condition. Based on our experience of Conditions 1 and 2, we proceeded to collect data for all conditions. As can be seen from Table 2, we dispatched a total of 3,013 requests to participate. About one-fourth of potential respondents (26%) opened the survey and, we presume, looked at the set-ups and questions. However, only about a third (or 9% of all potential respondents) actually completed the survey. The proportion 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 foot of Table 2). We will consider these outcomes again when we discuss the results below.

Insert Table 2 about here
----

[4] This was, in fact, done before the end of January 2010.

Table 2 documents characteristics of our respondents. In terms of position, a majority (59%) are at the rank of Associate Professor or higher. They also work in a wide variety of fields within the economics profession (respondents could indicate more than one area of specialization). Thirteen percent of respondents classified themselves as econometricians and more than two-thirds (77%) used regression analysis in their work (41% “often” or “always”). Whereas we cannot say whether our sample is representative of academic economists, it is quite large and undoubtedly captures a large number of competent professionals.

IV. Results

Condition 1

Respondents’ answers to Condition 1 are summarized in Figure 5. Three answers incorporating only “I don’t know”, or “?” were removed from the data. We also regarded as correct the answers of 4 participants who did not provide numerical responses, but mentioned that the answer was related to the error term and to its variance[5]. The questions and the correct answers are displayed in the titles of the histograms in Figure 5.

Insert Figure 5 about here

Most answers to the first three questions are incorrect. They suggest that the presentation of the results frames the respondents into evaluating the results only through the coefficient estimates and obscures the uncertainties implicit in the dependent variable. Specifically, Figures 5a through 5d show 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. A majority state that any small positive amount of *X* would be sufficient to obtain a positive outcome with 95% probability.
2. 68% of the answers to the second question suggest that for an individual to be better off with 95% probability than another person with $X=0$, a small amount of X ($X<10$) would be sufficient.

[5] Across all the Conditions there were 21 such responses.

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 of 51%).
4. 84% of participants gave a correct answer to question 4.

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 our sample of economists see as 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 and not the magnitude of the SDER. The apparent invisibility of the random component in the presentation seems to lure decision makers into disregarding the error term and to confuse an outcome with its estimated expected value. In their answers to questions 3 and 4, the majority of participants claim that if someone chooses $X=1$, the probability of obtaining $Y>1.001$ has a 50% chance, but obtaining $Y>0.936$ is almost certain.[6]

Our findings echo those of Lawrence and Makridakis (1989) who showed in an experiment that decision makers tend to construct confidence intervals of forecasts through estimated coefficients and fail to take into account correctly the randomness inherent in the process they are evaluating. They are also consistent with Goldstein and Taleb (2007) who have shown how failing to interpret a statistic appropriately can lead to incorrect assessments of risk. In the case of Condition 1, the information about the error term is not transparent; respondents only associate uncertainty with the coefficient estimates and their variation. This biases decision makers' perceptions of the predictability of outcomes.

In sum, the results of Condition 1 show that the common way of displaying results in the empirical economics literature obscures the uncertainty surrounding the analyzed outcomes. The data suggest that the lack of predictability is invisible to the

[6] Incidentally, the high rate of correct answers to question 4 suggests that 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").

respondents. In Condition 2, we tested this interpretation by seeing whether the answers to Condition 1 are robust to different levels of uncertainty.

Conditions 2 through 4

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

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

Insert Table 3 about here

Note that we are not arguing that economists consider only averages when making predictions. For most respondents, Conditions 1 and 2 ask unfamiliar questions that inquire about values and probabilities that are not related to coefficient estimates. They are not issues that are typically discussed in papers that publish regression analyses. Nonetheless, the similarity in responses between Conditions 1 and 2 shows that – under the influence of the current methodology – economists could be overestimating the effects of explanatory factors on economic outcomes. The misperceptions in the respondents' answers suggest that the way regression results are presented in publications can blind even the most knowledgeable individuals from differentiating among different clouds of data points and uncertainties. Parenthetically, at an early stage of our investigation, we conducted the same survey (using Conditions 1 and 2) with a group of 30 academic social scientists. The results (not reported here) were quite similar to those of our larger sample of economists.

Table 3 suggests that when the representation is augmented with a graph of actual observations and with statistical information on the magnitude of the error term (SDER), the perceptions of the relevant uncertainty and consequently the predictions improve.

However, around half of the participants still fail to take into account the error term when making predictions and give similar answers to those in Conditions 1 and 2. This suggests that respondents still mainly rely 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 (in the surveyed sample of studies only around 10% gave the SDER and 30% provided scatter plots with such detail), this does not provide much hope for improvement. Possibly more drastic changes in presentation are necessary to improve the perception of the predictability of an analyzed outcome. Conditions 5 and 6 were designed for this purpose.

Conditions 5 and 6

Our results so far suggest that, in trying to answer our questions, economists pay excessive attention to coefficient estimates and their standard errors and fail to consider the uncertainty inherent in the relation between the dependent and independent variables. What happens, therefore, when they cannot see estimates of coefficients and related statistics but only have a bivariate scatter plot? This is the essence of Conditions 5 and 6 – see the graphs in Figures 3 and 4.

Insert Figures 6 and 7 about here

Figures 6 and 7 display the histograms of responses to the four questions in Conditions 5 and 6, respectively. These show that participants are now much more accurate in their assessments of uncertainty compared to the previous Conditions (see also Table 3). In fact, when the coefficient estimates are not available, they are forced to pay attention solely to the graph, which depicts adequately the uncertainty within the dependent variable. This further suggests that scant attention was paid to the graphs when 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 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 information provided for the task. They claimed that, without the coefficient estimates, it was impossible to determine the answers and that all they did was to "guess" the outcomes approximately. Yet their guesses were more accurate than the predictions in the previous Conditions that resulted from careful investigation of the coefficient estimates and time-consuming computations. Indeed, as indicated in Table 2, respondents in Conditions 5 and 6 spent significantly less time on the task than those in Conditions 1 and 2 ($t = 2.95$ and 2.57, $p = 0.005$ and 0.01, respectively), and the participation rate was slightly higher although not statistically significant. In Appendix C, we provide a selection of comments made by respondents in all the conditions.

Effects of training and experience

Table 2 shows that our sample of 257 economists varied widely in terms of professorial rank, specialization within economics, and use of regression analysis in their work. Are these classifications related to accuracy in inferences made? Excluding Conditions 5 and 6 (where we have an unrealistic setting and answers were, on average, correct), we failed to find any relation between the numbers of correct answers and professorial rank or frequency of using regression analysis. On the other hand, a significantly higher percentage of statisticians, financial economists and econometricians performed well relative to the average respondent (with, respectively, 64%, 56%, and 51% providing correct answers compared to the overall average of 35%). When answers were accurate, the average time spent was also slightly higher, but the difference is not statistically significant (10.2 versus 9.3 minutes). Table 4 shows in detail the characteristics and proportions of respondents, who gave accurate answers in Conditions 1 through 4.

Insert Table 4 about here

V. Discussion

We conducted a survey designed to test the ability of economists to make probabilistic predictions from regression outputs presented in a manner similar to those published in leading economic journals. Given only the regression statistics usually reported in such journals, we find that many respondents made inappropriate inferences. In particular, they seemed to locate the uncertainty in prediction in estimates of the regression coefficients and not in the standard deviation of the estimated residuals (SDER). Indeed, responses hardly differed between cases where the "fit" of the estimated model varied between 0.25 and 0.50.

We also provided some of our respondents with scatter plots of the regression together with explicit information on the SDER. However, this had only a small ameliorative effect, which suggests that respondents relied principally on the regression statistics (e.g., coefficients and their standard errors) to make their judgments and did not make use of the information presented graphically. Finally, we forced other respondents to rely on graphical representation by only providing a scatter plot and no regression statistics. Interestingly, members of this group complained bitterly that they had insufficient information to answer the questions posed but, nonetheless, took less time to answer than the other groups and – most importantly – were more accurate in their responses.

The economists in our survey had various levels of seniority in the profession, specialized in different branches of economics, and made differential use of regression analysis in their work. Some of these characteristics were related to how they answered the questions we asked. In particular, whereas rank and frequency of regression usage were not related to respondents' performance, statisticians, financial economists and econometricians provided the most accurate answers to our questions.

Several objections could be made about our study in terms of, first: the nature of the questions asked; second, the particular respondents we managed to recruit; and third, the motivation of the latter to answer our questions.

First, we deliberately asked questions that are usually not posed in journal articles because we wanted to illuminate economists' appreciation of the predictability of economic relations as opposed to whether specific variables are or are not "significant." From a policy perspective, this is important (McCloskey & Ziliak, 1996). For example, even though economics articles typically do not address explicit decision making questions, economic models should be used to estimate, say, the probability of reaching given levels of output for specific levels of input. The questions are "tricky" only in the sense that they are not what economists typically ask. However, we always provided all the data needed to answer the questions correctly. Moreover, there was no time limit and respondents could use computational tools or textbooks if they wanted.

Second, we did not achieve a high response rate from our pool of respondents. As noted earlier, 26% of potential respondents took the time to open (and look at?) our survey questions and 9% answered. Does this mean, however, that our respondents were biased and, if so, in what direction? We don't know and would need to conduct further in-depth studies to find out. However, we did obtain a substantial number of respondents (257) who represent different characteristics of academic economists. Parenthetically, we wonder what kind of responses we might have received from those who opened the survey and then decided not to answer and particularly since in the "easiest" condition (only graphs) participants took the least time to answer (and were more accurate).

Third, by maintaining anonymity in responses, we were unable to offer incentives to our respondents. However, would incentives to answer these questions have made much difference? Clearly, without conducting a specific study we cannot say. However, extrapolating from results in experimental economics, the consensus seems to be that incentives increase effort and reduce variance in responses but do not necessarily increase average accuracy (Camerer & Hogarth, 1999). We also note that when professionals are asked questions relating to their competence, there would seem to be little incentive to provide a casual answer.

Parenthetically, it is possible to argue that our survey actually simulates quite well the circumstances under which many economists read journal articles: There are no explicit monetary incentives; readers do not wish to make additional computations; nor do they wish to do additional work to fill in gaps left by the authors; and time is precious. Thus, the framing of results by the authors is crucial.

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

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 results imply more 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 reservations. However, the typical form of presentation can lead to underestimating the uncertainty implicit in the underlying regression model. Specifically, there can be considerable variability around expectations of effects that 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 can be argued that the principle of *caveat emptor* should apply. That is, consumers of economic research should know better how to use the information provided and it is their responsibility if they underestimate the uncertainty implicit in the results they are examining. It is not the fault of the authors or the journals. We make two arguments against the *caveat emptor* principle as applied here.

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

Second, given the vast quantities of economic reports available today, 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 knowledge gained from economic research allows one to say. In short, they will believe that economic outputs are far more predictable than is in fact the case.

Parenthetically, we make all of the above statements assuming that econometric models describe empirical phenomena appropriately. At best, it can only be shown that model assumptions are approximately satisfied (they are not "rejected" by the data) and that, whereas the model-data fit is maximized within the particular sample observed, there is no guarantee that the estimated relations will be maintained in other samples. Indeed, the R^2 estimated on a fitting sample inevitably "shrinks" when predicting to a new sample and it is problematic to estimate *a priori* the amount of shrinkage.

Furthermore, in all the conditions, we assume that the errors are normally distributed, which might not be the case in naturally occurring settings. For instance, Taleb (2007) and Makridakis, Hogarth, and Gaba (2009) argue that statistical "outliers" are more common than typically assumed in social and economic environments and their effects are difficult to predict. Even though many estimation procedures in the published articles do not require normally distributed random disturbances to obtain consistent estimates, the explanations they provide through coefficient estimates and average values would be less accurate if the law of large numbers does not hold. Hence decisions that are weighted towards expected values and coefficient estimates would be even less accurate than our results indicate.

This discussion leads to considering what might be done to improve current practice. Our results show that providing graphs alone led to the most accurate inferences. However, since the comments made by our respondents were so negative about this format, we do not deem it to be a practical solution. On the other hand, we believe that it is appropriate to present graphs together with summary statistics as we did in Situations 3 and 4. However, given the inaccuracy of our respondents' answers, we believe that authors should provide aids in the form of internet links to sites that (a) explore different implications of the analysis, and (b) let readers pose different probabilistic questions. In short, we propose providing simulation tools that allow readers to experience the uncertainty in the outcomes of the regression.[7]

Whereas our suggestion imposes an additional burden on authors (which can be lower with experience), it reduces effort and misinterpretation on the part of readers, and

[7] For example, by following the link http://www.econ.upf.edu/~soyer/econometrics.html the reader can investigate many questions concerning the two regression set-ups that we examined in this paper as well as experience simulated outcomes (Soyer & Hogarth, in preparation).

makes the article a more accessible scientific product. Moreover, it has the potential to correct statistical misinterpretations that were not identified by our study. As such we believe our suggestion goes a long way to toward increasing understanding of economic phenomena. At the same time, our suggestion calls for additional research into understanding when and why different presentation formats lead to misinterpretation.

In addition to suggesting changes in how statistical results should be reported in journals, our results also have implications for the teaching of statistical techniques to economists. First, textbooks in econometrics should provide more coverage of how to report statistical results as well as instruction in how to make probabilistic predictions. Even a cursory examination of leading textbooks shows that these topics currently receive very little attention and provide incomplete. Indeed, the presentations provided in Conditions 1 and 2 are consistent with the prescriptions of Wooldridge (2008), the textbook that dedicates the largest space for such instructions, and thus presumably suitable for an empirical publication in economics. And yet, we have demonstrated that economists need more information to evaluate correctly the outputs of their analyses. Second, evaluating the predictive ability of economic models should become an important component of the teaching of econometrics. Indeed, if this is 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 predict and control different phenomena. However, if we cannot make appropriate probabilistic statements about our predictions, our ability to assess our knowledge accurately is seriously compromised.

References

Baltagi, B. H. (2007). Worldwide Econometrics Rankings: 1989 2005. *Econometric Theory*, Cambridge University Press, vol. 23(05), pages 952-1012, October.

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.

Gall, M. D., Borg, W. R., & Gall, J. P. (1996). *Educational research* (6th ed.). White Plains, NY: Longman.

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 edition. Upper Saddle River NJ: Prentice Hall.

Gujarati, D. N. (1988). *Basic Econometrics*. McGraw-Hill, New York.

Hogarth, R. M. (Ed.) (1982). *Question framing and response consistency: New directions for methodology of social and behavioral science*, No. 11. San Francisco, CA: Jossey-Bass.

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*. Wiley, New York.

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

King, G. (1990a). Stochastic Variation: A Comment on Lewis-Beck and Skalaban's 'The R-Squared'. *Political Analysis,* 2(1), 185-200.

King, G. (1990b). When Not to Use R-Squared. *The Political Methodologist,* 3(2), 11-12.

Krueger, J.S. & Lewis-Beck, M. S. (2007). Goodness-of-Fit: R-Squared, SEE and 'Best Practice'. *The Political Methodologist*, vol. 15, no. 1.

Lawrence, M. & Makridakis, S. (1989). Factors Affecting Judgmental Forecasts and Confidence Intervals. *Organizational Behavior and Human Decision Processes* 42, 172-187.

Lewis-Beck, M. S. & Skalaban, A. (1990a). The R-Squared: Some Straight Talk. *Political Analysis*, 2(1), 153-171.

Lewis-Beck, M. S. & Skalaban, A. (1990b). When to Use R-Squred. *The Political Methodologist,* 3(2), 9-11.

Lewis-Beck, M. S. & Skalaban, A. (1991). Goodness of Fit and Model Specification. *The Political Methodologist*, 4(1), 19-21

Makridakis, S., Hogarth, R. M., & Gaba, A. (2009). *Dance with chance: Making luck work for you.* Oxford, UK: Oneworld.

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

McNeil, B. J., Pauker, S. G., Sox, H. C., Jr., & Tversky, A. (1982). On the elicitation of preferences for alternative therapies. *New England Journal of Medicine*, 306, 1259-1262.

Pedhazur, E. J. (1997). *Multiple regression in behavioral research* (3rd ed.). Ft. Worth, TX: Harcourt Brace.

Soyer & Hogarth (in preparation). *When "experience" beats analysis: The advantage of simulation in statistical inference*. Barcelona, Universitat Pompeu Fabra.

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

Taleb, N. N. (2007). *The black swan: The impact of the highly improbable*. New York, NY: Random House.

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. (1986). Rational choice and the framing of decisions. *Journal of Business*, 59, S251-S278.

Willson, V. L. (1980). Research techniques in AERJ articles: 1969 to 1978. *Educational Researcher*, 9(6), 5-10.

Wooldridge, J. M. (2008). *Introductory Econometrics: A modern approach*. International Student Edition (3rd), Thomson, South Western.

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

Studies that	Journals: AER	QJE	JPE	RES	Total	% of 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%
...do NOT provide R^2 statistics	12	9	0	1	22	20%
...do NOT provide the sample standard deviation of the dependent variable(s)	21	9	4	0	34	31%
...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 SDER	5	3	1	1	10	9%

Table 2: Characteristics of respondents

Condition:	1	2	3	4	5	6	Total	%'s
Requests to participate	568	531	548	510	438	418	3,013	
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%
Senior Lecturer	0	2	1	0	0	2	5	2%
Lecturer	6	4	1	1	3	3	18	7%
Post-Doctoral Researcher	2	0	0	0	1	3	6	2%
Total	45	45	49	38	36	44	257	
Research fields								
Econometrics	14	11	10	14	6	12	67	13%
Labor economics	12	11	14	10	9	8	64	12%
Monetary economics	5	2	5	2	2	1	17	3%
Financial economics	4	5	4	3	1	1	18	3%
Behavioral economics	3	7	2	3	2	6	23	4%
Developmental economics	8	2	9	5	5	2	31	6%
Health economics	4	3	5	1	1	4	18	3%
Political economy	3	5	7	4	3	4	26	5%
Public economics	9	6	10	8	4	6	43	8%
Environmental economics	1	2	3	2	1	1	10	2%
Industrial organization	2	6	6	2	8	7	31	6%
Game theory	4	1	4	5	2	7	23	4%
International economics	6	6	7	2	2	3	26	5%
Macroeconomics	9	9	13	6	6	6	49	9%
Microeconomics	11	4	11	7	9	10	52	10%
Economic history	2	2	6	2	1	2	15	3%
Statistics	3	6	1	1	0	3	14	3%
Other	0	0	1	0	0	0	1	0%
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 on survey	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 through 6

Condition:	1	2	3	4	5	6
R^2	*0.50*	*0.25*	*0.50*	*0.25*	*0.50*	*0.25*
Percentage of participants whose answer to:						
Question (1) was X < 10 *(Incorrect)*	**72**	**67**	**61**	**41**	**3**	**7**
Question (2) was X < 10 *(Incorrect)*	**68**	**70**	**67**	**47**	**3**	**15**
Question (3) was above 80% *(Incorrect)*	**60**	**63**	**63**	**49**	**9**	**7**
Question (4) was approx. 50% *(Correct)*	**84**	**88**	**93**	**84**	**91**	**93**
Approximate correct answers are						
Question 1	48	84	48	84	48	84
Question 2	67	118	67	118	67	118
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	43	49	37	32	43
Question 4	44	41	49	37	32	43

Notes:

(1) For questions 1, 2, and 3, there are significant differences between Conditions 1 and 5, 3 and 5, 2 and 4, and 4 and 6 ($t > 2.90$, $p < .01$, for all comparisons).

(2) For question 4, there are no significant differences between Conditions.

(3) There are no significant differences between Conditions 1 and 2, 3 and 4, and 5 and 6 for all questions.

Table 4: Relations between training, experience and responses in Conditions 1 to 4 (number of respondents with correct answers in parentheses)

Condition	1	2	3	4	Total over four conditions	Proportion 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

Figure 1: Presentation of Condition 1

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	1000

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.

Figure 2: Tables in Condition 2 (the rest of the presentation is the same as Figure 1)

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

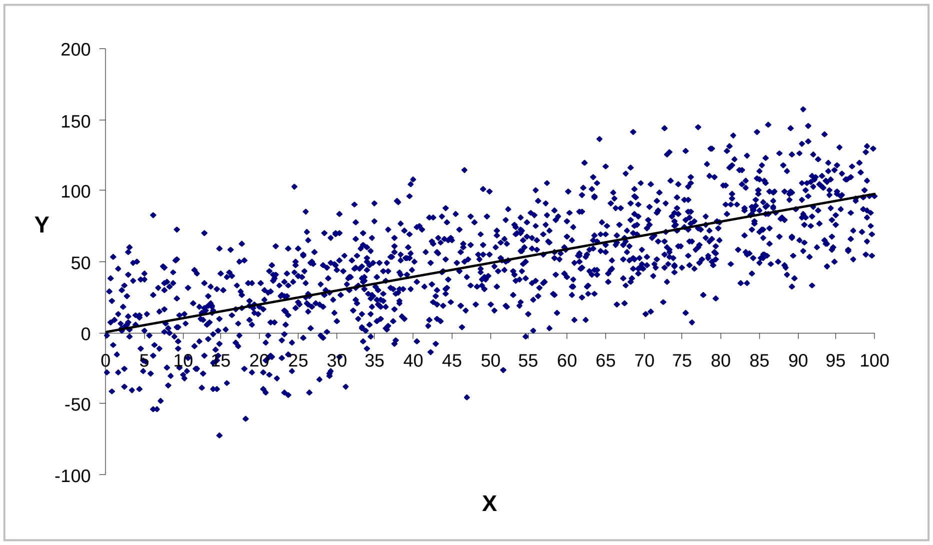
	Dependent Variable: *Y*
X	1.02 (0.056)**
Constant	0.61 (3.74)
R^2	0.25
N	1 000

Standard errors in parentheses

** Significant at 95% confidence level

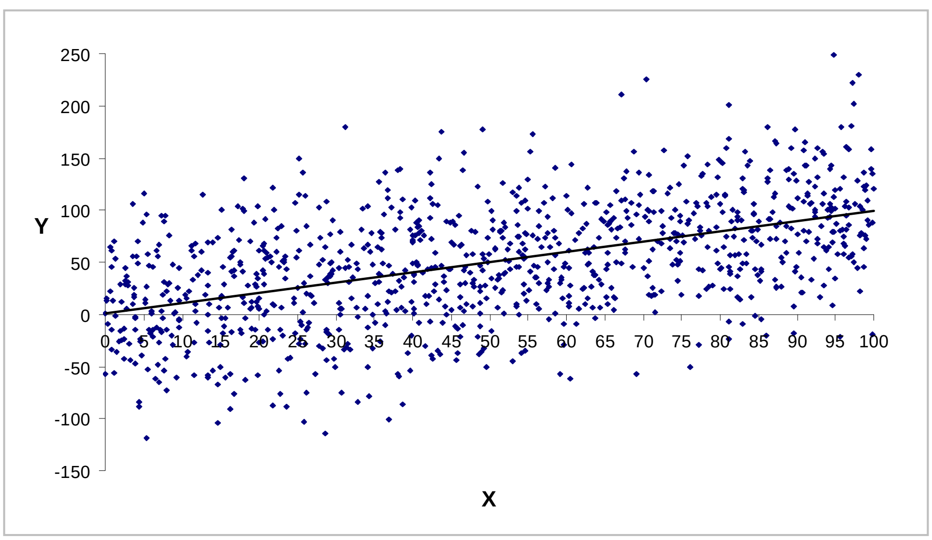
N is the number of observations

Figure 3: Bivariate scatter plot of Condition 1 and information on SDER



Note: The standard deviation of the estimated residuals is 29.

Figure 4: Bivariate scatter plot of Condition 2 and information on SDER



Note: The standard deviation of the estimated residuals is 51.

Figure 5: Histograms for the answers to Condition 1

Figure 5a: Answers to (1) in Condition 1 *(N=39)*

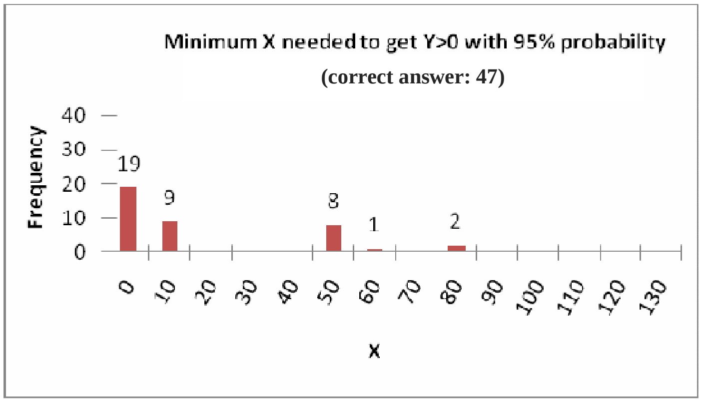


Figure 5b: Answers to (2) in Condition 1 *(N=35)*

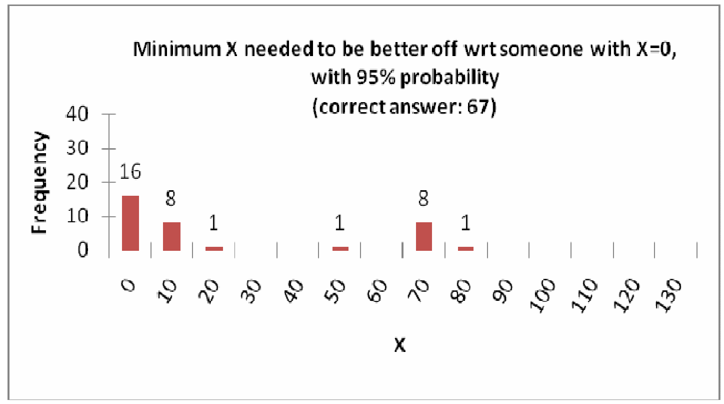


Figure 5c: Answers to (3) in Condition 1 *(N=45)*

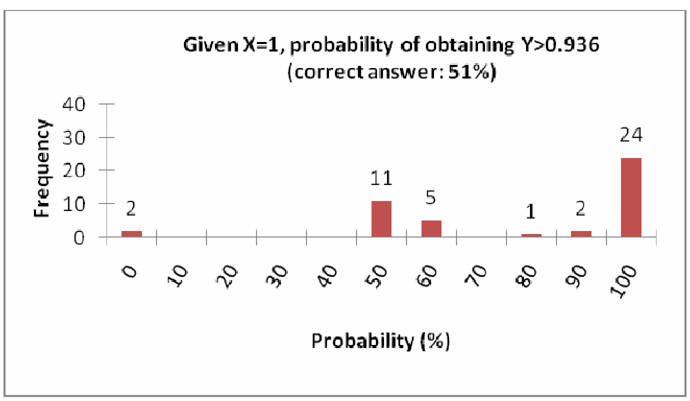
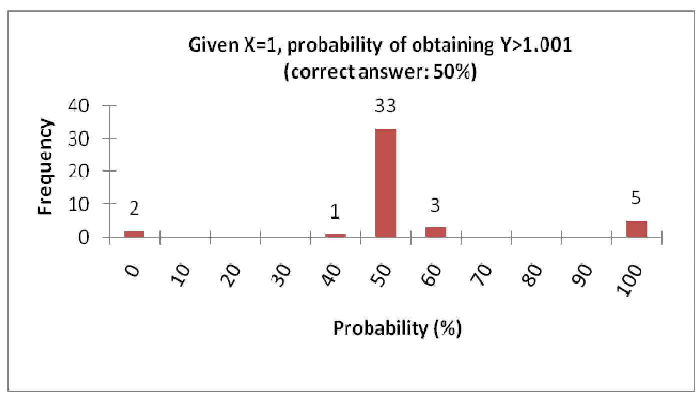


Figure 5d: Answers to (4) in Condition 1 *(N=44)*



<u>Figure 6: Histograms for the answers to Condition 5</u>

Figure 6a: Answers to (1) in Condition 5 *(N=31)*

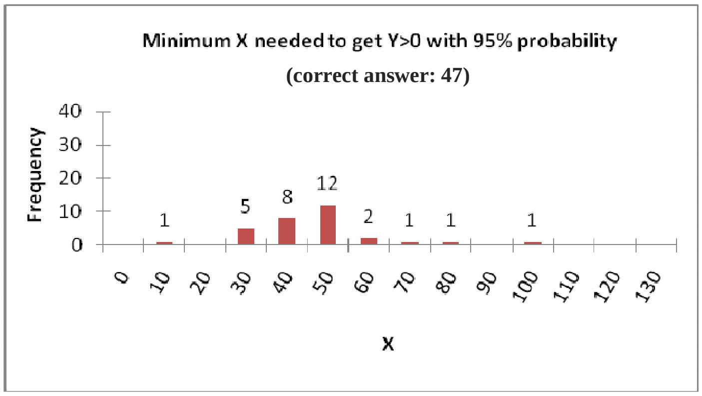


Figure 6b: Answers to (2) in Condition 5 *(N=30)*

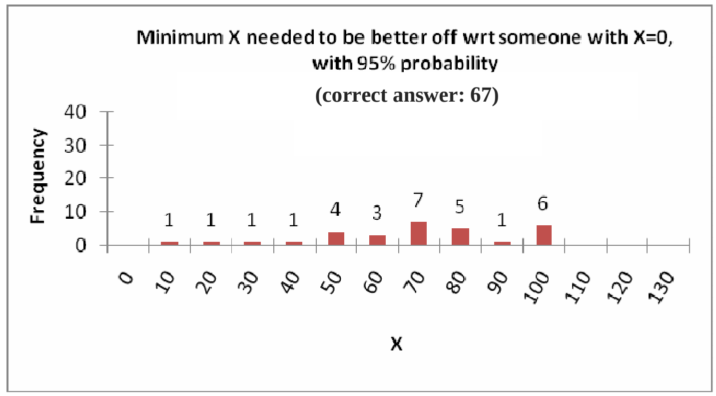


Figure 6c: Answers to (2) in Condition 5 *(N=32)*

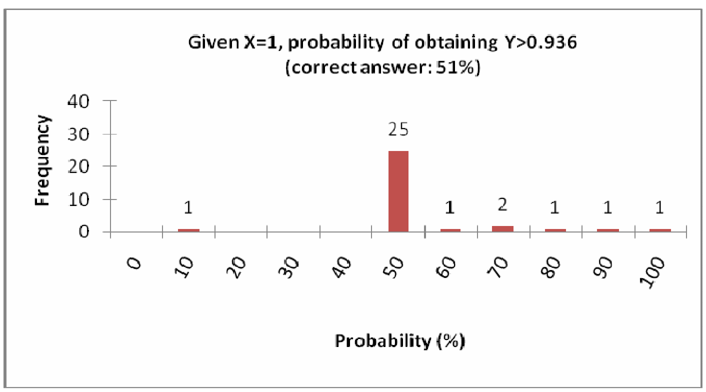
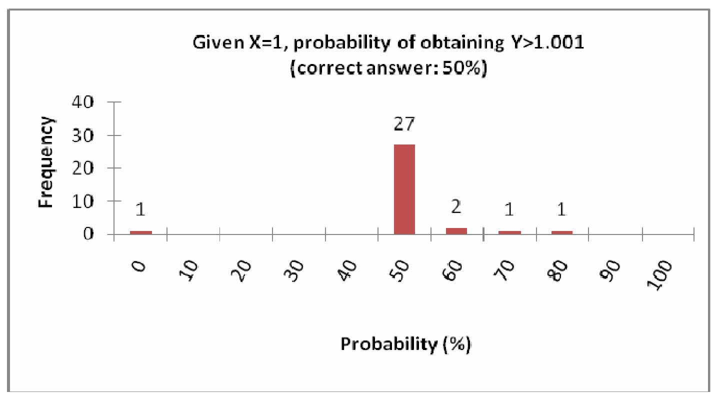


Figure 6d: Answers to (4) in Condition 5 *(N=32)*



<u>Figure 7: Histograms for the answers to Condition 6</u>

Figure 7a: Answers to (1) in Condition 6 *(N=41)*

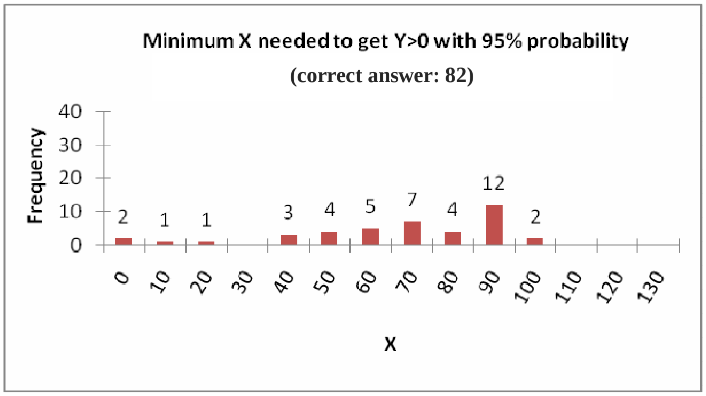


Figure 7b: Answers to (2) in Condition 6 *(N=39)*

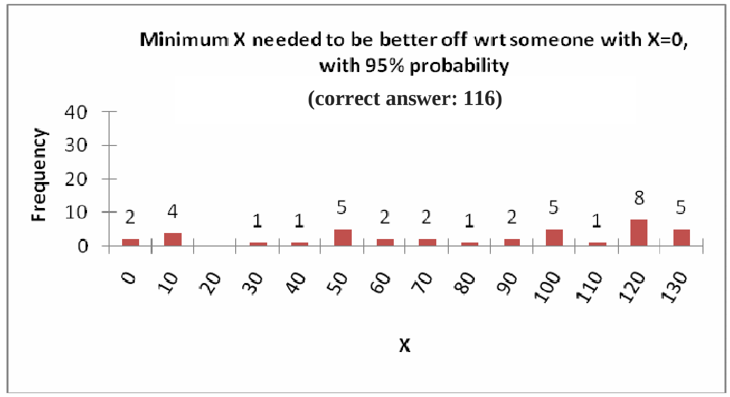


Figure 7c: Answers to (3) in Condition 6 *(N=43)*

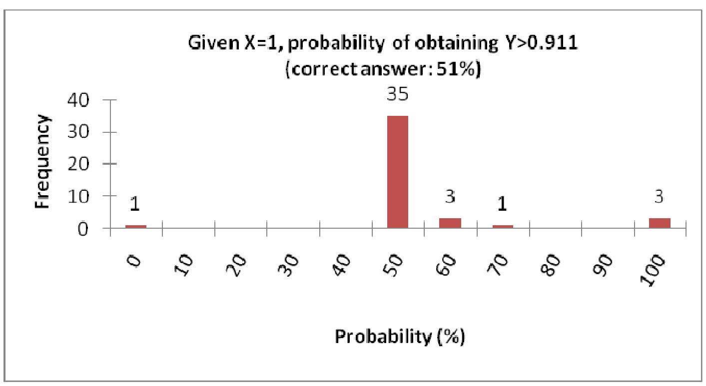
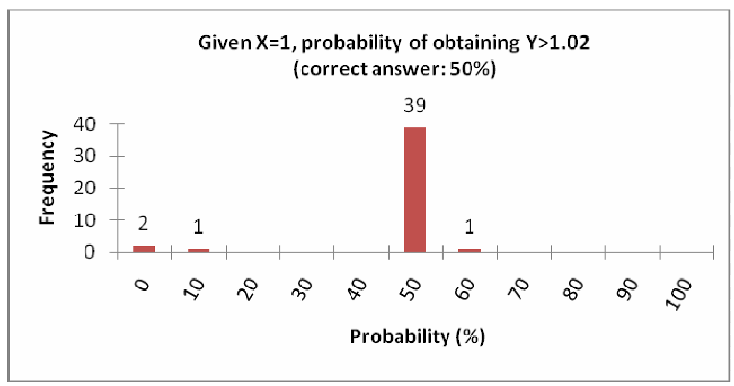


Figure 7d: Answers to (4) in Condition 6 *(N=43)*



<u>Appendix A: Rationale for answers to the four questions</u>

Preliminary comments

We test whether or not decision makers knowledgeable about regression analysis correctly evaluate the unpredictability of an outcome, 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*. Furthermore 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 the assumptions of the classical linear regression model hold (see, e.g., Greene, 2003).

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 relation 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 not 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 (SDER).

In a linear regression analysis, SDER^2 corresponds to the variance of the dependent variable that is unexplained by the independent variables and is captured by the statistic $(1\text{-}R^2)$. In Conditions 1 and 3 this is given as 50%. One can compute the SDER using the $(1\text{-}R^2)$ statistic and the variance of *Y*:

$$\text{SDER} = se(\hat{e}) = \sqrt{(Var(Y)(1-R^2)} = \sqrt{(40.78^2)(0.5)} \cong 29 \qquad \text{(A1)}$$

The answer to the first question can be found by constructing a one-sided 95% confidence interval using (A1). We are looking for X where;

$$\text{Prob}(Z > -(\hat{C}+\hat{\beta}X)/se(\hat{e})) = \text{Prob}(Z > -(0.32+1.001X)/29) = 0.95 \text{ ; } Z \sim N(0,1) \qquad \text{(A2)}$$

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

$$X = (1.645 * 29 - 0.32)/1.001 \cong 47 \qquad \text{(A3)}$$

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

$$(Y_i \mid X_i = x_i) - (Y_j \mid X_j = 0), \text{ where } x_i > 0. \qquad \text{(A4)}$$

We know that $(Y_i \mid X_i)$ is an identically, independently and normally distributed random error with an estimated standard deviation of again 29. Given that a different and independent shock occurs for different individuals and actions, the standard deviation of (A4) becomes:

$$\begin{aligned}&\sqrt{Var\left[(Y_i \mid X_i = x_i) - (Y_j \mid X_j = 0)\right]} \\ &= \sqrt{Var(Y_i \mid X_i = x_i) + Var(Y_j \mid X_j = 0)} = \sqrt{(29^2 + 29^2)} \cong 41 \qquad \text{(A5)}\end{aligned}$$

Thus, the answer to question 2 is:

$$X = (1.645 * 41 - 0.32)/1.001 \cong 67 \qquad \text{(A6)}$$

For Condition 2 (and thus also 4 and 6), similar reasoning is involved. For these conditions, the equivalent of equation (A1) is

$$\text{SDER}= se(\hat{e}) = \sqrt{(Var(Y)(1-R^2)} = \sqrt{(59.25^2)(0.75)} \cong 51 \qquad \text{(A7)}$$

such that the answer to question 1 is:

$$X = (1.645 * 51 - 0.62)/1.02 \cong 82 \qquad \text{(A8)}$$

As for question 2:

$$\sqrt{Var\left[(Y_i \mid X_i = x_i) - (Y_j \mid X_j = 0)\right]} = \sqrt{(51^2 + 51^2)} \cong 72 \qquad \text{(A9)}$$

So that the answer to question 2 is:

$$X = (1.645 * 72 - 0.62)/1.02 \cong 116 \qquad \text{(A10)}$$

Answers to Questions 3 and 4

Here, we inquire about how decision makers weight the different sources of uncertainty within the dependent variable, given the typical communication of estimation results.

These questions provide insight as to whether or not the presentation of the results frames the participants into considering 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 SDER from equation (A1), that is

$$\begin{aligned}
&\Pr(\mathrm{Y_i} > 0.936 \mid \mathrm{X_i} = 1) = \Pr(\hat{\mathrm{C}} + \hat{\beta}\mathrm{X_i} + \hat{\mathrm{e}} > 0.936 \mid \mathrm{X} = 1) = \\
&= \Pr(\hat{\mathrm{e}} > 0.936 - \hat{\mathrm{C}} - \hat{\beta}\mathrm{X} \mid \mathrm{X} = 1) = \Pr(\frac{\hat{\mathrm{e}}}{\mathrm{se}(\hat{\mathrm{e}})} > \frac{0.936 - \hat{\mathrm{C}} - \hat{\beta}\mathrm{X}}{\mathrm{se}(\hat{\mathrm{e}})} \mid \mathrm{X} = 1) = \qquad \text{(A11)} \\
&= 1 - \Phi(\frac{0.936 - 0.32 - 1.001}{29}) = 1 - \Phi(-0.013) \cong 0.51
\end{aligned}$$

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 to (A11) in order to obtain an answer.

$$\Pr(Y_i > 1.001 \mid X_i = 1) = \Pr(\hat{C} + \hat{\beta}X_i + \hat{e} > 1.001 \mid X = 1) =$$
$$= \Pr(\hat{e} > 1.001 - \hat{C} - \hat{\beta}X \mid X = 1) = \Pr(\frac{\hat{e}}{\mathrm{Var(e)}} > \frac{1.001 - \hat{C} - \hat{\beta}X}{\mathrm{Var(e)}} \mid X = 1) =$$
$$= 1 - \Phi(\frac{1.001 - 0.32 - 1.001}{29}) = 1 - \Phi(-0.01) \cong 0.5 \tag{A12}$$

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

$$\Pr(Y_i > 0.911 \mid X_i = 1) = \Pr(\hat{C} + \hat{\beta}X_i + \hat{e} > 0.911 \mid X = 1) =$$
$$= \Pr(\hat{e} > 0.911 - \hat{C} - \hat{\beta}X \mid X = 1) = \Pr(\frac{\hat{e}}{\mathrm{Var(e)}} > \frac{0.911 - \hat{C} - \hat{\beta}X}{\mathrm{Var(e)}} \mid X = 1) = \tag{A13}$$
$$= 1 - \Phi(\frac{0.911 - 0.61 - 1.02}{51}) = 1 - \Phi(-0.015) \cong 0.51$$

And for question 4,

$$\Pr(Y_i > 1.02 \mid X_i = 1) = \Pr(\hat{C} + \hat{\beta}X_i + \hat{e} > 1.02 \mid X = 1) =$$
$$= \Pr(\hat{e} > 1.02 - \hat{C} - \hat{\beta}X \mid X = 1) = \Pr(\frac{\hat{e}}{\mathrm{Var(e)}} > \frac{1.02 - \hat{C} - \hat{\beta}X}{\mathrm{Var(e)}} \mid X = 1) = \tag{A14}$$
$$= 1 - \Phi(\frac{1.02 - 0.61 - 1.02}{51}) = 1 - \Phi(-0.01) \cong 0.5$$

<u>Appendix B: Histograms for the answers to Conditions 2, 3 and 4</u>

Condition 2

Figure B2a: Answers to (1) in Condition 2 *(N=36)*

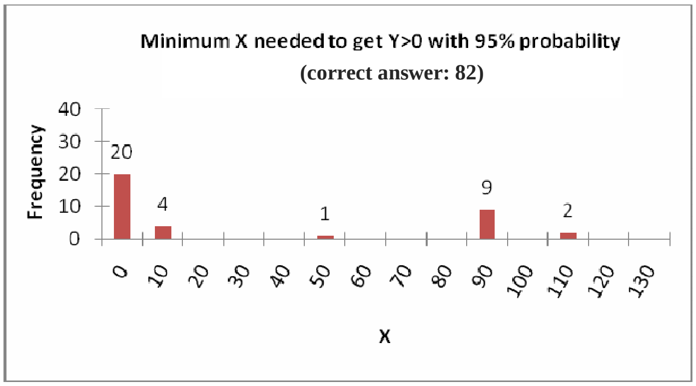


Figure B2b: Answers to (2) in Condition 2 *(N=30)*

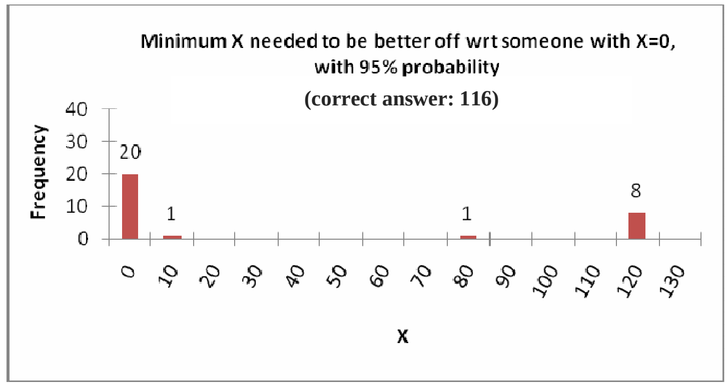


Figure B2c: Answers to (3) in Condition 2 *(N=43)*

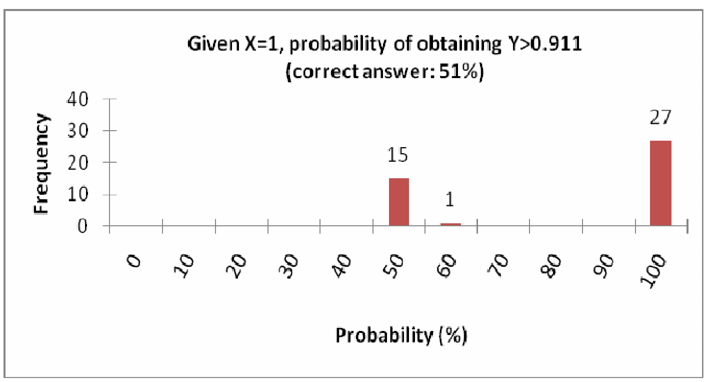
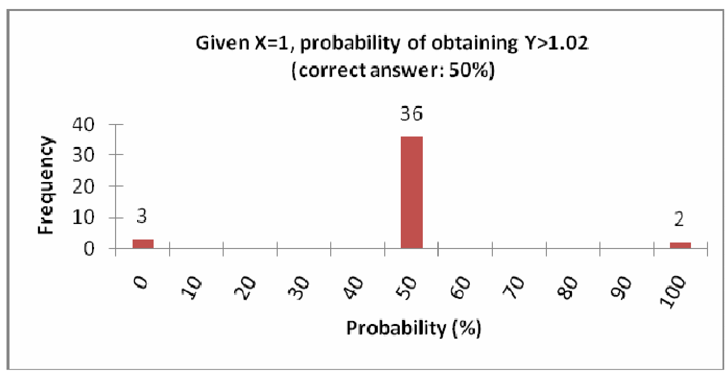


Figure B2d: Answers to (4) in Condition 2 *(N=41)*



<u>Condition 3</u>

Figure B3a: Answers to (1) in Condition 3 *(N=44)*

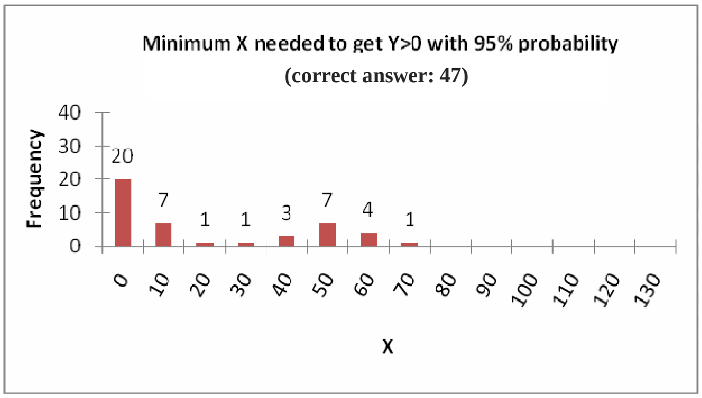


Figure B3b: Answers to (2) in Condition 3 *(N=39)*

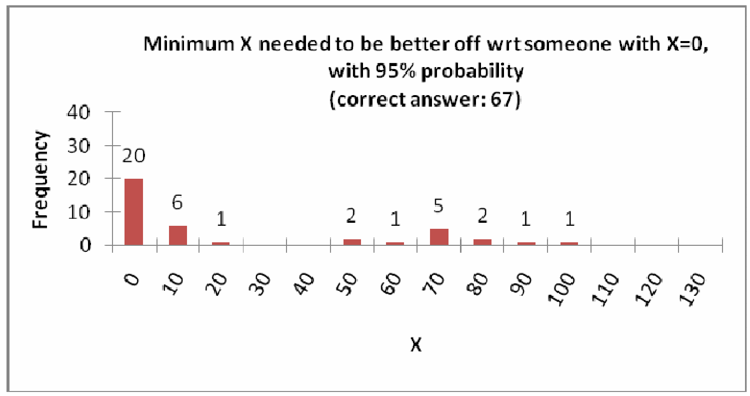


Figure B3c: Answers to (3) in Condition 3 *(N=48)*

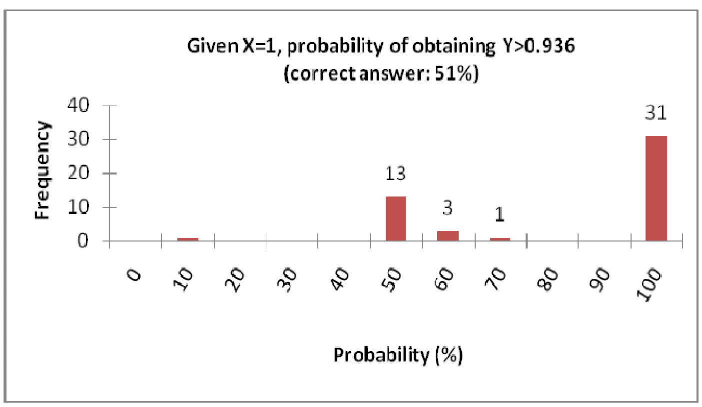
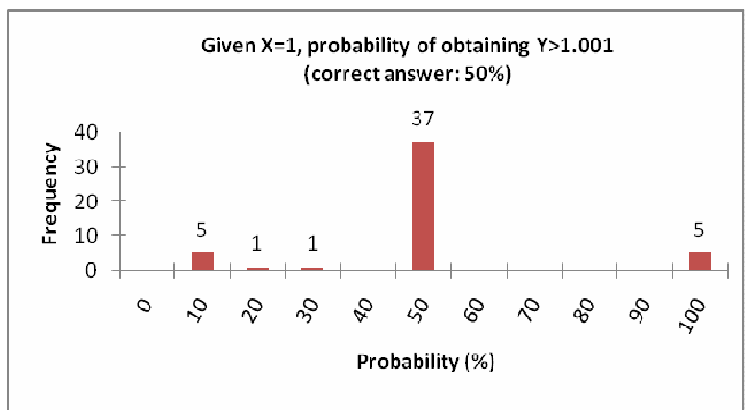


Figure B3d: Answers to (4) in Condition 3 *(N=49)*



Condition 4

Figure B4a: Answers to (1) in Condition 4 *(N=32)*

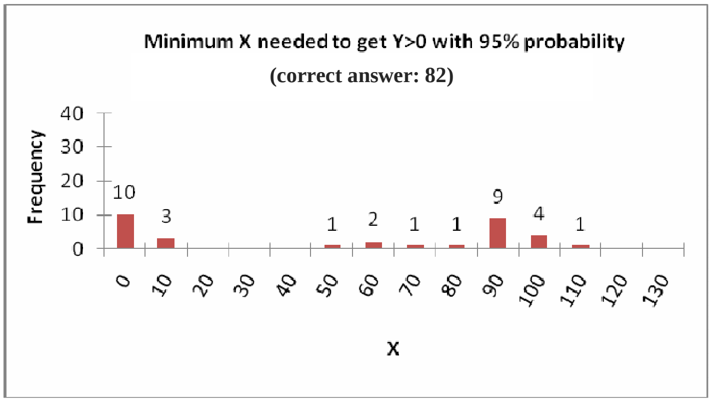


Figure B4b: Answers to (2) in Condition 4 *(N=32)*

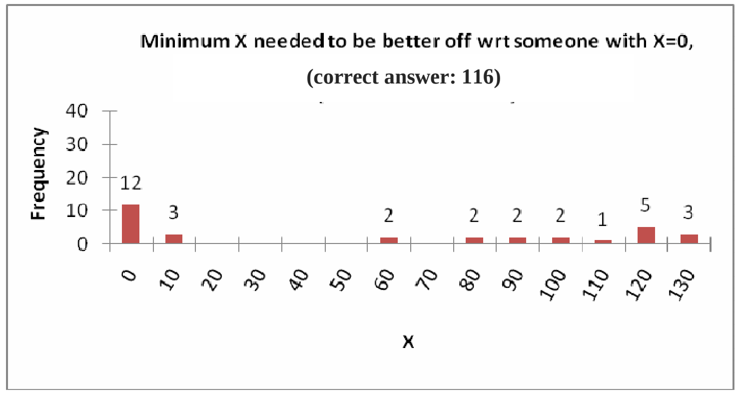


Figure B4c: Answers to (3) in Condition 4 *(N=38)*

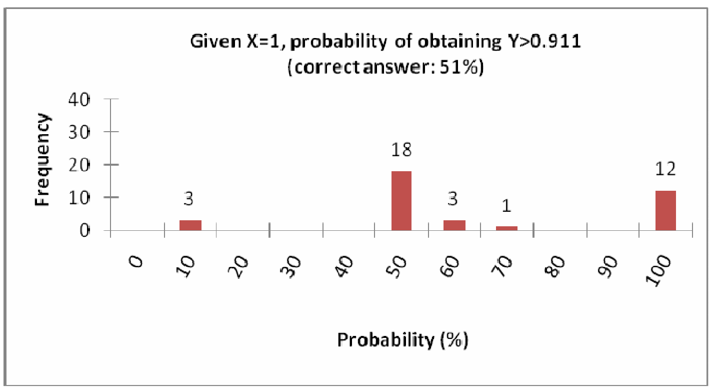
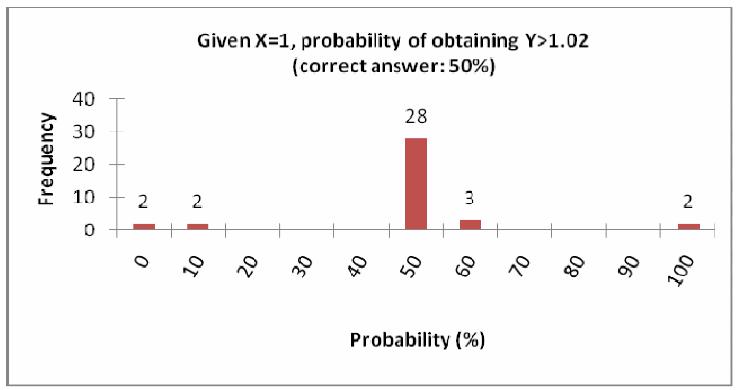


Figure B4d: Answers to (4) in Condition 4 *(N=37)*



Appendix C: Some of the comments made by the participants

Conditions 1 and 2

A) Selected comments made by those who answered correctly.

- This was surprisingly hard. It made me realize how little attention I generally pay to the constant, which is so important for statements about predicted values.
- Not sure what you expect from this survey - Most people who estimate regressions in economics don't use them for prediction but for making inferences on the values of the coefficients.
- A graph would help!
- Your first question with the X=1 is a bit odd given average X=50. Would be nice if after the done button we get the answers.
- I may have misunderstood the question, but the regression results don't easily allow you to answer it. It seems to require a forecast and standard error of forecast for a value of X which is far from the mean. It would require some calculation to get these from the reported results.
- Your questions come across as trick questions, though I'm guessing that they are not supposed to be.

B) Selected comments made by those who answered incorrectly.

- I don't think the questions you have in this survey are the typical questions an empirical researcher is interested in as far as regression results are concerned. This looks more like a set of simple statistics exercises. Anyway, it was fun. Good luck with your research!
- Given the assumptions laid out at beginning, we know that Yhat = Chat+Bhat*X is an unbiased estimate of the true Y. But to answer the questions, we need to know the variance of the prediction error, that is Var(Y-Yhat). This variance is: Var(Y-Yhat) = Var(Yhat) + Var(u), where we can estimate Var(u) by SSR/(n-1) = [59.25^2*999*(1-0.25)]/999, and where Var(Yhat) depends on Cov(Chat,Bhat), which is not provided. Given the assumptions, this error will be approximately normally distributed.

- I believe these statements have not been expressed in a statistically correct way; it should refer to confidence levels instead of probabilities.

C) Selected comments made by those who did not answer.

- If this is a true experimental context with respect to X, model appears not correctly specified due to constant. Q1-Q4 can be answered using standard software so why would anyone do calculations by hand.

Conditions 3 and 4

A) Selected comments made by those who answered correctly.

- Trying to think carefully through these questions, I realized that I would have to review the theory of forecast error, which is not something that I use regularly. Perhaps the answers are easy in the end, but one would have to review quite a bit of regression analysis to verify that point to oneself.
- I just made quasi-educated guesses. I'm curious to see how wrong or right they are.
- Could you send to the participants to this survey what your main results are? It would be simply matter to send a link to the same mailing list used to ask for compiling the survey. Your topics (though quite uncommon for the typical quantitative analysis) are very interesting. My compliments.
- I enjoyed taking the survey. I would be interested in learning about the results.
- Very nice questions!
- To first order, only the variation in the error term matters here. Also, I only did the calculations very approximately. I think the graph is helpful; it reminded me that the root MSE is important here. Nice question to ask.
- I used plot more than numbers for this
- Cute. I bet you get lots of wrong answers!
- All probabilities are computed conditionally on the parameter estimates, i.e. conditionally on the observed sample
- Definitely an interesting question. I would appreciate it if you could send me the precise answers.

- I used the scatter plot to draw my inferences rather than the estimated relationship. I would be interested to know how accurate my answers were (or the correct answers for each value of x). Thanks for including me.

B) Selected comments made by those who answered incorrectly.

- These questions are phrased in very unusual ways, ways that are much different than when I teach the concepts in undergraduate econometrics.
- It was irritating
- the questions are too messy and not clear enough
- These questions appear to suppose that the constant can be set to zero. If one does this, the estimated slope coefficient will change.
- You may want to also provide the covariance between the estimates of C and beta.

- I found the way the questions are posed unclear, in that the preamble says "the model links Y and X correctly" -- I find that ambiguous: are we supposed to assume that, or are you telling us that is true?

C) Selected comments made by those who did not answer.

- Why should I do this?
- I do not wish to spend the time working these out. However, I would get the bayesian predictive density for Y, which would be a t-distribution, and work out the probability from that. Other answers might (1) treat the betahats as equal to the true values and ignore uncertainty in estimation of the betas, or (2) treat the betas as uncertain with st deviations given by the st errors of the betahats. I view neither of these as correct. The Bayesian solution is the only coherent one.

Conditions 5 and 6

A) Selected comments made by those who answered correctly.

- I did not use any statistical knowledge to arrive at my answers, simply looked at the diagram.
- The scatter plot was not really helpful to draw inferences, so I was guessing from visual inspection.

- Was I supposed to have estimates for the coefficients, coefficient standard errors, and error term (I didn't)? All I could really do was eyeball the graph and take a guess at the coefficients and proceed from there.
- The questions seem to be ones that are not often addressed based on regression analysis, and this particular presentation did not include information that would be very useful in getting the answers and is normally generated by regression analysis.
- Interesting problem. I hope I answered correctly!
- This was fun!
- I've answer the previous questions under the assumption that the 1000 data where generated by the econometric model. I don't see how to estimate beta with the information provided, then gave a look to the graph.
- I am not quite sure what you are getting at here, but I do not think the questions are well posed. For a start, you should have given the fitted regression equation and assoc. standard errors. Also, the first two questions are virtually indistinguishable from one another (other than to state that the second probability is slightly less than the first). Obviously I do not know what your precise motive is here, but I am not sure you have thought this out carefully enough to achieve it! Good luck in your work.

B) Selected comments made by those who answered incorrectly.

- The assumptions on the model are incomplete. In order to answer these questions you need to compute conditional probabilities and for this you need to assume conditional normality (or joint normality of e and X). Note that normality of e alone does not allow you to compute the required probabilities.

C) Selected comments made by those who did not answer.

- Not enough info.
- I would need to know the standard error of the beta coefficient to answer.