Visualizing Uncertainty

On Soyer's and Hogarth's "The illusion of predictability:
How regression statistics mislead experts"

By Stephen T. Ziliak

Trustee and Professor of Economics
Roosevelt University
430 S. Michigan Ave
Chicago, IL 60605 USA
Email: sziliak@roosevelt.edu
http://sites.roosevelt.edu/sziliak
http://stephentziliak.com

July 2012

Abstract: This comment was published in the *International Journal of Forecasting* symposium on the Soyer-Hogarth experiment (Vol. 28, No. 3, July/Sept. 2012, pp. 712-714). The experiment evaluates the ability of expert econometricians to make predictions based on commonly provided regression output. Visual displays of quantitative information, including simple plots of data, outperformed predictions based on R-squared, t-statistics, and other common diagnostics. Reliance on graphing - on the visualization of uncertainty - was suggested more than a century ago by Karl Pearson, a founding father of English language statistics. The results of the Soyer and Hogarth experiment, when combined with evidence produced by Ziliak and McCloskey (2008) and others, suggests that graphing and visualization should receive more attention and tests of statistical significance, less.

Keywords: regression analysis, forecasting, uncertainty, visualization, statistical significance, Karl Pearson

JEL Classifications: C1, C18, B1

Visualizing Uncertainty
On Soyer's and Hogarth's "The illusion of predictability: How regression statistics mislead experts"

By Stephen T. Ziliak[1]

A great Bayesian statistician, decision theorist, and mathematical economist, Leonard "Jimmie" Savage (1917-1971), said famously that the best test of a hypothesis is provided not by the statistical significance of coefficients nor by fits of R-squared but instead by "interocular" evidence that "hits you between the eyes"—for example, in a scatter plot of data graphed together with a theoretical prediction equation or regression line (Edwards, Lindman, & Savage, 1963; cf. Zellner, 2004). Savage—though a Ph.D. in theoretical mathematics—nevertheless believed that the visual display of quantitative information—or what Savage called the "interocular trauma test"—is the best way for an econometrician to see a real and important difference, if any, between the model and the data. Savage is hardly alone but in English language statistics the graphical school has not been much heeded.

Perhaps it should be. Soyer & Hogarth (2011) performed an experiment on the behavior of well-published econometricians, testing their ability to predict economic outcomes using the conventional outputs of linear regression analysis: standard errors, t-statistics, and R-squared. The chief finding of the Soyer-Hogarth experiment is that the leading econometricians themselves—our best number crunchers—make better predictions when *only* graphical information—such as a scatter plot and theoretical linear regression line—is provided to them. Give them t-statistics and fits of R-squared for the same data and regression model and their forecasting ability declines. Give them *only* t-statistics and fits of R-squared and predictions fall from bad to worse.

It's a finding that hits you between the eyes, or should. R-squared, the primary indicator of model fit, and t-statistics, the primary indicator of coefficient fit, is in the leading journals of economics (for example, in the AER, QJE, JPE, and RES, examined by Soyer & Hogarth) doing more harm than good. Soyer & Hogarth find that conventional presentation mode actually damages inferences from models, and thereby hurts decision making by reducing the econometrician's (and profit seeker's) understanding of the total error of the experiment—or of what might be called the real standard error of the regression, where "real" is defined as the sum (in percentage terms, say) of both systematic and random sources of uncertainty in *the whole model*.

[1] Stephen T. Ziliak is Trustee and Professor of Economics, Roosevelt University. He thanks J. Scott Armstrong, Andrew Trigg, and James P. Ziliak for helpful comments. His email address is: sziliak@roosevelt.edu.

Lacking a graph to visualize total model uncertainty, econometricians lack a good method for predicting. Shy of a basic scatter plot, econometricians do as they were trained to do by journal referees and by the vast majority of graduate level textbooks: they focus attention on R-squared and t-statistics, and thus vastly underestimate—or vastly overestimate—the total level of uncertainty (Ziliak & McCloskey, 2008).

I wish that Soyer & Hogarth had gotten a better than 9% response rate. I wish they would have discussed the representativeness of the 9% that actually completed the anonymous questionnaire. And I wish their survey of best journal practice included articles on time series econometrics: it did not (but see: Klein, 1997). Still, these minor complaints do not subtract from the truth of their chief finding and claim. My feeling is that the problem documented by Soyer & Hogarth is in fact much worse, that economists' prediction ability is far worse than is evidenced by Soyer & Hogarth by virtue of the fact that the authors examined only econometricians who publish in leading econometrics journals.

Indeed it was John Maynard Keynes (1937, p. 214), the founder of modern macroeconomics and the author of an important book on probability theory, who acknowledged variables for which "there is no scientific basis on which to form any calculable probability whatever. We simply do not know." Examples of variables we do not know well enough to forecast include "the obsolescence of a new invention", "the price of copper" and "the rate of interest twenty years hence" (Keynes, p. 214). More recently Taleb (2007, this issue) observes that to calculate probability one can certainly turn to dice throws, and many—despite Keynes and the recently "unexpected" Great Recession—do. But probability-based forecasting is not valid if the dice thrower does not know how many sides appear on the die or which corners have been shaved flat or rounded off completely. The probabilistic search for external validity—the quest for knowledge of general variance and the out-of-sample experience of life—is itself a dice throw however scarcely acknowledged.

Statistics was graphical at its formal inception more than a century ago. Not so much now, a fact which is, Soyer & Hogarth find, causing bad decisions to be made and good decisions to be made by accident or, more frequently, not at all.

Can a graphical display of data enable better analysis? It seems so, and some of the best minds of statistics agree. A half century before Savage invoked his "interocular trauma test" Karl Pearson (1857-1936) established the first department of statistics (which he called "applied statistics"), at University College London, in 1911. Yet War-time budgets were tight and his fledgling department—despite the undoubted greatness of the founder himself—sputtered along until the late 1920s, when he was given enough resources to replace G. Udny Yule (who moved on to government work and a job at Cambridge) with

Jerzy Neyman (from Poland) and Pearson's own son, Egon S. Pearson (from Cambridge), for example, who helped to make UCL the leading site for statistical inquiry in the world.

As editor of *Biometrika* for more than 32 years, Karl Pearson introduced research workers to correlation, regression, the normal distribution, the chi-square distribution, skew distributions, biometry, large sample statistical significance testing, and graphing (Porter, 2004; Ziliak & McCloskey, 2008, chps. 18-19).

Yet few realize that Pearson himself—the inventor of, among many other things, the "histogram" employed by Soyer & Hogarth—believed that statistics should be considered a subset of a more general field, *graphics*—or what Pearson sometimes called "graphical statics" and even "the geometry of statistics" (Porter, 2004, pp. 236-237).

Pearson promoted the graphical approach to statistics in lectures he gave at University College London and elsewhere. Said Yule: "The first course [taught by Pearson in 1894 and attended by Yule] opened with a brief outline sketch of history, leading up to a 'Kollektivmass' definition of statistics. . . . but would any other lecturer have thought of suggesting the study of Marey's *La method graphique dans les sciences experimentales* (1875, 1885)? Karl Pearson," said Yule, "was an enthusiast for graphic representation and thought in graphic terms" (Yule quoted in Pearson, 1990, p. 14).

Siding with a French tradition against the then dominant German, Pearson was inspired in part by engineers and by, for example, Etienne-Jules Marey, a French physiologist who is credited with innovating a number of visual tools, including modern cinematography. "Geometry is not merely a mode for representing research," Pearson said, "*but it is essentially a mode of statistical research….Most statistical conclusions which can be obtained by arithmetic, can be obtained also by geometry, and many conclusions can be formed which it would be difficult to reach except by geometry*" (Pearson quoted by Porter, 2004, p. 236-237; emphasis in original).

"He was also familiar with the growing use of graphs, usually time series, in English empirical economics" (Porter, p. 236), his biographer notes, and above all he seemed to be inspired most by the economist Alfred Marshall's essay "On the Graphic Method of Statistics," which Marshall (1885)—to whom we owe the modern graph of supply and demand (Gordon, 1982)—published in the Jubilee Volume of the Royal Statistical Society—ironically, in the same volume that published the first use of the word "significance" in a statistical testing context, as in "statistical significance" (Edgeworth, 1885; see Ziliak & McCloskey, 2008, chp. 17, for discussion).

By the late 1920s Pearson's large sample biometrics was being replaced by small sample analysis pioneered by William Sealy Gosset (1876-1937) aka "Student" in his job as experimental brewer of Guinness (Ziliak, 2008). Gosset attended Pearson's statistics lectures during the academic year 1906-1907, while Gosset was on paid sabbatical at University College London. Gosset himself was a visualizer, and he made many of his most important points—in reports at the brewery, in articles in *Biometrika*, and in letters of correspondence with other statisticians—using graphical methods (Ziliak & McCloskey, 2008, chp. 20). Gosset's quality control charts were picked up and advanced by Shewart and Deming, where they continue to influence presentation mode in operations research (for example, Student, 1927).

Thus Soyer & Hogarth are mistaken when they assert that the origin of today's presentation mode—of non-graphical statistics—is not known or well understood. It is.

The practice of reporting standard errors, t-statistics, and R-squared begins in best-selling books by Ronald A. Fisher (1890-1962), who—as I have shown in a number of articles and in a book with Deirdre N. McCloskey—directed research attention away from evaluation of magnitudes of coefficients and of the real standard error of the regression—the very quantities which the forecaster and decision maker need to know—and toward the quantities that, as Soyer & Hogarth, like Ziliak & McCloskey before them show, don't actually work.

By the 1940s the Cowles Commission approach to econometrics—despite resistance by Frisch, Haavelmo, and Tinbergen—embraced Fisherian statistics. Thus the multiple-equation macro-structural forecasting models by Lawrence Klein, for example, typify the non-graphical approach which came to characterize late 20th century statistics and today's presentation mode.

What types of graphics work best Soyer & Hogarth do not say and evidently did not test (cf. Tufte, 2001). But the practice and theory of graphing is as old as Babylon, and can be tested. In economics, for example, in the late 19th century, the founding fathers of neoclassical economics (Alfred Marshall and others) argued passionately about the proper role, value, and meaning of graphical display. Now that Soyer & Hogarth have pointed out the cost to forecasters and decision makers of non-graphical display, perhaps they and others will advance their work by testing the accuracy of prediction by type and style of graphical display. After all, artists, astronomers, psychologists, and physicists have conducted many types of experiments on presentation mode: see, for example, Hecht, Schwartz, and Atherton, eds., (2003). And the role of visualization in the construction of scientific knowledge and persuasion has recently caught the attention of evolutionary biologists and cognitive scientists as well as historians and philosophers of economics and statistics (for example: Giere, 1996; Klein,1997). Additional experimental studies of the Soyer-Hogarth type will add to the growing stock of knowledge.

With extremely powerful, dynamic, and high-dimensional visualization software programs such as “GGobi”, which is currently being provided to users for free on-line, economists can join engineers, cancer researchers, and rocket scientists, and do a lot more gazing at data than we currently do (http://www.ggobi.org). At least, that is, if our goal is to improve decisions and to identify relationships that hit us between the eyes.

References

Edgeworth, F. (1885). Methods of statistics. *Jubilee Volume of the Statistical Society*, 181–217. London: Royal Statistical Society of Britain.

Edwards, W., Lindman, H., & Savage, L.J. (1963). Bayesian statistical inference for psychological research. *Psychological Review*, 70, 193-242.

Giere, R. (1996). Visual models and scientific judgment. In: *Picturing knowledge: Historical and philosophical problems concerning the use of art in science*, Toronto: University of Toronto Press. B. S. Baigrie, (Ed.), 369-402.

Gordon, S. (1982). Why did marshall transpose the axes? *Eastern Economic Journal*, 8, 31-45.

Hecht, H., Schwartz, R. & Atherton, M. (Ed.). (2008). *Looking into pictures: An interdisciplinary approach to pictorial space*. Cambridge: Cambridge University Press.

Keynes, J. (1937). The General Theory of Employment. *Quarterly Journal of Economics*, 51, 209-223.

Klein, J. (1997). *Statistical visions in time: A history of time series analysis from 1662 to 1938.* Cambridge: Cambridge University Press.

Marey, E.J. (1885). *La methode graphique dans les sciences experimentales*. Paris: Masson. Second printing.

Marshall, A. (1885). On the graphic method of statistics. *Jubilee Volume of the Statistical Society*, 251–260. London: Royal Statistical Society of Britain.

Pearson, E. (1990) [posthumous]. *'Student': A statistical biography of william sealy gosset*. Oxford: Clarendon Press. R. Plackett, (Ed.), with the assistance of G. Barnard.

Porter, T. (2004). *Karl pearson: The scientific life in a statistical age*. Princeton: Princeton University Press.

Soyer, E. & Hogarth, R. (2011). Presentation mode in econometrics and the illusion of predictability. *International Journal of Forecasting,* this issue.

Student. (1927). Errors of routine analysis. *Biometrika*,19, 151-64.

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

Tufte, E. (2001). *The visual display of quantitative information.* Cheshire, CT: Graphics Press. Second edition.

Zellner, A. (2004). *Statistics, econometrics, and forecasting.* Cambridge: Cambridge University Press.

Ziliak, S. (2008). Guinnessometrics: The economic foundation of 'Student's' t. *Journal of Economic Perspectives*, 22, 199-216.

Ziliak, S. & McCloskey, D. (2008). *The cult of statistical significance: How the standard error costs us jobs, justice, and lives.* Ann Arbor: University of Michigan Press.