Dusek, Tamás (2009) 'A review of Stephen T. Ziliak and Deirdre N. McCloskey, *The Cult of Statistical Significance. How the Standard Error Costs Us Jobs, Justice, and Lives*', *The Journal of Philosophical Economics*, II:2, 121-124

A review of Stephen T. Ziliak and Deirdre N. McCloskey, *The Cult of Statistical Significance. How the Standard Error Costs Us Jobs, Justice, and Lives*, The University of Michigan Press, Ann Arbor, 2008, 320 pages

Tamás Dusek

The problems of statistics do not mainly concern the mathematical structure of the method; they are generally about the adequate application of mathematical theory to an observed real world phenomenon. The most striking misuse of a statistical method in the field of economics and many other social and behavioral sciences is perhaps the highly corrupt practice of inferential statistic and statistical significance test. The book under review gives an excellent overview of the methodological, historical, institutional and personal-psychological aspects of the misuse of statistical significance test in various branches of science and offers a solution for the untenable situation.

The structure of the work can be briefly described as follows. After the preface and introductory chapter, the first four chapters deal with fundamental concepts of significance test, such as Type I and Type II error, loss function, distinction between scientific and statistical significance and the importance of sample size. Chapters 5-10, mainly based on previous works of the authors, comprise a discussion of the practice of economics concerning significance test. Chapters 11-16 are devoted to presenting the practice of psychology and medicine from the same point of view. Detailed general information about the historical origin of significance test is provided in Chapters 17-23. Chapter 24 plans the solution. This structure is supplemented by the permanent confrontation of Gosset's (aka Student of Student's t-test) method and Fisher's method on significance test.

Statistical significance is a measure of sampling variance which "is sometimes interesting, but a low value of it is not the same thing as scientific importance. Economic significance is the chief scientific issue in economic sciences". (p. 249) According to Ziliak and McCloskey, statistical significance tests are widely

misused in many fields of applied science. The textbooks and teaching are wrong, the most prestigious journals follow wrong practices concerning articles, which contain statistical significance tests. This critique is not originated in antiquantitative or antistatistician attitude, but quite contrary, criticizing a wrong practice the authors defend the sound statistical method. "Statistics, magnitudes, coefficients are essential scientific tools. No one can credibly doubt that. And mathematical statistics is a glorious social and practical and aesthetic achievement. (...) Our book is not a tract against counting or statistics. On the contrary. In our own scientific work we are quantitative economists and value statistics as a crucial tool." (p. 1) Moreover, Ziliak and McCloskey claim that "scientific assertions should be confronted quantitatively with the world as it is or else the assertions is a philosophical or mathematical one, meritorious no doubt in its own terms but not scientific." (p. 4) I do not share this method-oriented demarcation of science and philosophy.

The majority of the most important statements are introduced in the first chapter. Sizeless scientists treat statistical significance as scientific significance and statistical insignificance as scientific insignificance. The judgment about scientific or substantive significance is not possible according to standardized, mechanistic rules but it requires thought, creativity and imagination. The Fisherian way of making a significance test is criticized, that is, concentrating only on the chance of finding a difference where there is none (Type I error). Science needs loss function, it is important to take into consideration the chance and the cost of finding no difference when there is one (Type II error). As regards the costs of the bad practice of statistical significance tests, the concrete examples of the book originate not from economics, but mainly from health sciences. The examples of economics illustrate the cost of an inadequate method in a more general and indirect way: the manipulation with probability distributions and significance tests is an unjustifiable and disturbing part of the results of analysis of economic data.

I focus only on the discussion of the practice of significance test in economics. The authors examined according to 19 viewpoints the practice of applying significance test of the articles in the American Economic Review. While the concerned questions are important, many of them are relevant only in those cases, where the data comes from either random samples, or from a repeatable experiment or from an observation of a random (stochastic) process. Then the significance test informs us that the observed connection between two phenomena can occur randomly or

not at some level of probability. Ziliak and McCloskey do not deal in detail with the more elementary epistemological problems: "the application of sampling theory to non-samples such as entire populations, or, more controversially, the application of sampling theory to entire instantiations of a time series. (...) Suppose for the moment that these other problems have been solved or at any rate set aside." (p. 66) The sixth question refers to the problem of sample or population: "Does the article refrain from reporting t- or F-statistics or standard errors even when a test of significance is not relevant?" (p. 68) Less than 10 percent of the articles are correct from this point of view, both in 1980s and 1990s.

I think the troubles rooted in this type of deeper epistemological problems and some of those factors which are examined by the authors (for example examining the power function, considering the power of the test) are important only in those cases where the underlying epistemological assumptions fit. Unfortunately this is not typical in economics or econometrics. While macroeconomic data are clearly not random samples, they are treated often as if it would be. Macroeconomic indices are aggregated (through time, space, individuals, quality and behavior), weighted and, contrary to physics, biometrics and psychometrics, not results of designed experiments. Macroeconomic indices are not one actualisation of repeatable 'random samples' derived from a larger population but a part of economic history. The most influential paper about the use of probabilistic models in economics is Haavelmo's paper, "The Probability Approach in Econometrics" (1944). Haavelmoo's paper does not argue that economic data are random samples, but it asserted that economic phenomena have stochastic characteristic. Mathematical economists and scholars of econometrics till Haavelmo made deterministic mathematical models. Havelmoo's main argument for probabilistic approach is the following: it is well known that there isn't an exact functional relationship between observable economic variables. Actual observations will deviate more or less from any exact functional relationship. And, according to Haavelmo if some relationship is not exact, then it is stochastic. This strange opinion became widely and rapidly accepted by the majority of the new generation of economists. The main problem of this statement is that it is grounded on a false dichotomy between stochastic and deterministic processes. It is not true that if some relationship is not exact or deterministic then it has to be stochastic or probabilistic. It can be uncertain also, without any numerical probability. And we can recognize the uncertain character of a relationship not by examining the numbers themselves but by examining the qualitative information about the data generator process.

Dusek, Tamás (2009) 'A review of Stephen T. Ziliak and Deirdre N. McCloskey, *The Cult of Statistical Significance. How the Standard Error Costs Us Jobs, Justice, and Lives*', *The Journal of Philosophical Economics*, II:2, 121-124

The difference between the deterministic and stochastic phenomena is epistemologically by far not as important as the fundamental differences between stochastic and uncertain phenomena. Stochastic and uncertain phenomena have nothing in common apart from the incompleteness of our knowledge. For example, the result of a sport competition is clearly not deterministic. If it was deterministic, the results would be known in advance. If someone thought, objective numerical probability can be given to the various possible results, then she or he could consider the sport competition as a stochastic phenomenon. But I think this would be a metaphysical, mystical and unjustifiable treatment of the point, because the alleged numerical probabilities are not verifiable. Every sport competition is a unique phenomenon; the circumstances are different in each competition. We can express only in numerical terms also our subjective opinion about the outcome of events. As a more complex example for uncertain phenomena in the field of economics, it can be predicted qualitatively that the increase of money supply leads to the increase of price level in the unspecified future, but it is impossible to predict both the timing of the process in exact numerical terms and the exact effect of the increasing money supply to the structure of the price system, to income distribution, to the rate of interest, to the change of production structure and so on. If someone still gave a prediction in quantitative form about the process – and this is a common practice of economic policy research institutions – this prediction cannot be treated as a precise and exact numerical result neither in a deterministic nor stochastic sense, but as an indicator of direction and magnitude of the examined process.

As a conclusion the authors write: "Economic scientists – for example, those who submit articles to the AER or edit or referee it or some other journal or serve on hiring and grant-making committees – routinely violate elementary standards of statistical cogency. (...) This should cease. The economics profession should set meaningful standards of economic significance." (p. 121) The last chapter of the book is about "What to Do". The first suggestion concerns the education of statistics, the second the current practice of statistical analysis. The book should offer stimulating reading for those who are interested in the sound practice of statistics and hopefully make a significant impact on reducing the misuse and overemphasis of statistical significance test in every field of research.

Tamás Dusek is Associate Professor at Széchenyi István University, Győr, Hungary (dusekt@sze.hu)