

# On the Method of Social Sciences and Statistics

Matteo Barigozzi \*

April 2007

PRELIMINARY IDEAS PLEASE DO NOT QUOTE  
COMMENTS ARE WELCOME

On 25<sup>th</sup> March 1938, Ettore Majorana, an Italian physicist and one of Enrico Fermi's colleagues, mysteriously disappeared during a journey from Palermo to Napoli. During his life he only published nine papers, all about physics. Some of them constituted a fundamental contribution to the development of quantum mechanics. A tenth paper was published thanks to Majorana's brother and Giovannino Gentile in 1942. An English translation of the same paper was published in 2005 with the title "The value of statistical laws in physics and social sciences".

Here I want to review the main points of the paper by Majorana that contains some very up to date considerations on the widely used (and abused) methodological parallelism between physics and economics. I will also add some personal opinions on the road that I think should be followed by empirical economics.

Mechanics is the most exact branch of physics, it expresses all natural laws as deterministic laws in a pure mathematical way. When Majorana was active it was believed that also economics should be formulated as a deterministic theory with its own concept of equilibrium. Walrasian economics is the typical example of such a theory. What happens when in physics we consider complex systems of particles such as gases? The laws governing the particles are deterministic, therefore we could in principle solve simultaneously all the equations of motion, one for each particle. It turns out to be more feasible and useful to consider the aggregated properties of the gas which can be obtained using the methods of statistics. Already in the field of classical mechanics, statistics is a useful tool when we want to study systems of many particles. The main statistical concept used in physics is the Law of Large Numbers which applies to independent and identically distributed quantities, i.e. to homogenous particles. The output is always a Gaussian distribution (e.g. in the kinetic theory of gases by James Clerk Maxwell). In neoclassical economics agents are homogeneous therefore the Law of Large Numbers can be applied. The use of the representative agent in this case is similar to considering physical aggregates instead of single particles. However, Majorana was among the first to discover the laws of quantum mechanics in the 1930s. According to this new physics,

---

\*Sant'Anna School of Advanced Studies - Laboratory of Economics and Management  
Piazza Martiri della Libertà 33, I-56127 Pisa, Italy.

*E-mail:* matteo.barigozzi@gmail.com

the fundamental laws are not deterministic but are instead stochastic. The use of statistics to study the properties of all microscopic systems becomes then unavoidable. Statistics is used again to aggregate, but it is used also to represent the laws of motion of single particles. More recent developments of quantum mechanics (e.g. by Richard Phillips Feynman) tend to express again the laws of physics as perturbations of deterministic laws, thus repositing somehow the deterministic character of nature and the notion of equilibrium. Statistics is however still essential if we want to study aggregated properties of quantum complex systems and the Law of Large Number is still one of its main tools.

According to Majorana, the revolution of quantum mechanics brings physics close to economics, indeed the laws governing the social systems at microscopic level seem to be more stochastic than deterministic. The value of statistics is two-folded: it is needed to describe the behaviour of the single agent and to aggregate the behaviours of many agents into macroscopic variables. Concerning the second point, in economics we have a more complex situation than in physics: particles are homogeneous, while individuals are heterogeneous and often the Law of Large Number cannot be applied. As in physics, the aggregation of many stochastic laws generates quantities that follow deterministic paths and on which the policy maker can intervene.

The questions I want to challenge are: what makes social sciences governed by stochastic laws? How do we describe and study social systems? Is there a science with methods that maybe suitable for economics? How do we aggregate microscopic heterogeneous behaviours?

We can recognize at least two kinds of sources of stochasticity in human behaviour: one is given by the uncertain environment in which decisions are taken, while the other is more linked to human perceptions. Concerning the former, we remember first of all the contribute of Carl Menger in 1871, according to whom the value depends on the subjective evaluations of individuals, on their needs and on the way they try to satisfy them. Moreover these evaluations may change according to the environment and in an unpredictable way. The limited knowledge of individuals makes the decision making process uncertain, and the concept of equilibrium becomes vague and difficult to apply. The other and more important contribute on limited knowledge comes from the book “A treatise on probability” written by John Maynard Keynes in 1921. For him individuals are rational, but lack a defined opinion. They can only attach their degree of confidence to events given their knowledge. Together with this measure, Keynes divides the events between those for which a probability measure can be computed (short-period uncertainties as when choosing production levels) and those for which the probability is unknown (long-term uncertainties as when choosing investment levels). These latter events are determinant for the general theory of Keynes. The well known “animal spirits” represent the residual uncertainty which is inevitable in subjective evaluations. This form of stochasticity that we find in social sciences can be formulated in terms of omitted variables, i.e. if individuals would have perfect information they would behave exactly in the same rational manner.<sup>1</sup> The lack of perfect information limits the extent to which agents can make a fully rational decision, thus, as Herbert Simon in the 1970s recognized, they possess only “bounded rationality” and must make decisions by “satisficing” or choosing not what is optimal but what makes them happy enough.

---

<sup>1</sup>Aside notice that this is not the case for quantum mechanics for which an interpretation in terms of omitted or hidden variables was confuted by the physicist John Stewart Bell in 1987. His theorem states that “no physical theory of local hidden variables can ever reproduce all of the predictions of quantum mechanics”.

A second kind of uncertainty in human behaviour is linked to psychology. Prospect theory was introduced by Daniel Kahneman and Amos Tversky in the 1970s and 1980s, after conducting experiments to study decision-making under risk. The main features of the theory are that people tend to think of possible outcomes usually relative to a certain reference point (often the status quo) rather than to the final status, a phenomenon which is called framing effect. Moreover, they have different risk attitudes towards gains (i.e. outcomes above the reference point) and losses (i.e. outcomes below the reference point) and generally care more about potential losses than potential gains (loss aversion). Finally, people tend to overweight extreme but unlikely events, but they underweight average events. This framework turned out to be useful to interpret many behaviours that seemed to deviate from standard economic rationality, one for all was the case of the equity-premium puzzle.

Heterogeneity of agents, limited knowledge, bounded rationality and even non-rational behaviours, together with the processes of learning by individuals and selection by the environment are the main features of decision-making in economics. It is very hard to aggregate such a microscopic complex system into a simple macroscopic system relying on the concept of equilibrium. After all in physics particles are homogeneous and the laws that govern their motion, although stochastic, are well defined. Thus a deterministic stable law at a macroscopic level is a legitimate formulation of the nature. Physics, even in its quantum formulation, is too deterministic, while this is not the case for economics where the evolutionary aspect of social systems and the heterogeneity of their constituents are the key features. Biology seems to be a more appropriate methodological paradigm to follow than physics. Moreover, biology is a young science as economics and still relies more on empirical observations, experiments and descriptions of the facts than on analytical formalization of the observed phenomena. The tool required by biology is statistics more than mathematics. We need the capacity of dealing with real data, of interpreting the results of the experiments, and of aggregating microscopic behaviours in macroscopic quantities, which are easier to observe. After a huge amount of data is collected and elaborated so that we have a sufficient number of robust stylized facts, then a theory can be formulated using these findings as hypothesis.

It is worth to remember some empirical approaches in economics. The most extreme atheoretical approach to social sciences comes from the German historical school of the 1850s (e.g. Bruno Hildebrand and Karl Knies). They consider economics as an essentially empirical science, refusing a priori hypothesis and deductive reasoning until it is reached a degree of knowledge of the reality sufficient to build the general hypothesis from which a theory can be built. In the same cultural environment Ernst Engel published in 1857 an empirical study on the consumption structure for different income levels. The Engel's law is still one of the most robust statistical regularities found in microeconomics. Finally, among the revisionists of Marx's thought, Sidney and Beatrice Webb founded in 1895 the London School of Economics with the objective of "the scientific and objective discovery of facts".<sup>2</sup> Arthur Bowley was one of the first scholars at the London School of Economics and it is known for his empirical approach to economic issues and for the publication in 1901 of the first book of statistics in English called "Elements of Statistics".

What happens when Menger tries to propose an analytical approach to build an evolutionary theory as suggested by the observation? For him the main assumptions of the theory come from direct experience, but when trying to formulate an analytical theory he cannot escape

---

<sup>2</sup><http://www.lse.ac.uk/resources/LSEHistory/lse.htm>

the concept of equilibrium. The same problem of a conflict between concepts and formalization is found in the work by Alfred Marshall. His focus is on the dynamical processes and on the evolutionary aspects of economic development, but he remains trapped by the concept of equilibrium when trying to formalize his ideas in a rigorous manner. A step forward in this direction was made by the evolutionary models by Richard Nelson and Sidney Winter in the 1980s. The recent advances in evolutionary modelling were possible thanks to improved mathematical techniques derived from the study of stochastic processes in statistical mechanics and in physics in general.

In a synthesis between the pure historical-inductive observation-driven (historicist) and the abstract-deductive logic-driven (marginalist) approaches, Keynes refused in all his works the use of general deterministic models, while he always preferred models aimed at explaining single, less general problems. This is often done by using simple non-analytic schemes of hierarchical relations of causes and effects to represent the relations between macroeconomic variables, which are generated by individual decisions taken in an uncertain environment. This aspect of Keynes' theory was greatly ignored by his followers, surviving only in the Post-Keynesian school. Joseph Schumpeter, in one of his first works in 1908, agreed with Keynes on the impossibility of formulating exact laws. In his writings he recognizes the need for the theory to be verified by the empirical reality. In particular first of all he states that the assumptions should give reasonable results. However, although this point might remind Milton Friedman's "as if" concept, Schumpeter adds also the requirement that hypothesis should be formulated always after considering the facts.

What is the role of econometrics? In 1944, Trygve Haavelmo proposed a probability based approach to economic models. The models used in economics do not make precise predictions that allow rejection of the theory based on precise measurements as instead we would like for every science. If models are to represent science, they must be confronted with data. However, in economics there is almost never the possibility of creating random experiments, i.e. that can be repeated as many times as it is necessary to reduce standard errors and to lead to robust conclusions. What Haavelmo proposes is to use stochastic equations where a systematic part is separated from the error on which a probability distribution is explicitly put. The lack of data coming from random experiments can be filled by considering past observations as many replications of an experiment. The need of studying time series is then clear. The approach by Haavelmo was also useful in providing a reply to the critique of Keynes to Jan Tinbergen concerning the use of econometric models for validating a theory. In fact econometric analysis of time series data could result in formal tests of theories. What is required is that the theories can be formulated so as to yield probability models for the observed data. In this respect a Bayesian approach seems more flexible than the standard econometric approach, in that it assigns probabilities also on the parameters of the model and not only on the data.

The problem of the non-availability of data obtained from random experiments is what distinguishes economic from biology and natural sciences in general. With the words of Christopher Sims:

"Economists can do very little experimentation to produce crucial data. This is particularly true of macroeconomics. Important policy questions demand opinions from economic experts from month to month, regardless of whether professional consensus has emerged on the questions. As a result, economists normally find themselves considering many theories and models with legitimate claims to

matching the data and to predicting the effects of policy. We have to deliver recommendations, or accurate description of the nature of the uncertainty about the consequences of alternative policies, despite the lack of a single accepted theory. Because non-economists often favor one policy or another based on their own interests, or prefer economic advice that pretends to certainty, there is an incentive for economists to become contending advocates of theories, rather than cool assessors of the state of knowledge”.

This is why the statistical approach is crucial for economists, because uncertainty cannot be reduced by replicating experiments.<sup>3</sup> Some sciences like physics do not need a statistical approach as they can be formulated as deterministic theories. Sciences like biology use statistics to collect data, but can reduce uncertainty by means of experiments. Finally other sciences like anthropology do not even consider the possibility of using statistical inference, but simply rely on the narrative description of facts. This latter case typically happens when no theory is available as an a priori guide. Indeed the typical economic situation is that of theories that are neither true nor false when confronted with real data. However, statistics is useful also in describing phenomena without any a priori theory. As Sims suggests, we should consider an atheoretical approach where the structure of the model is not predetermined (e.g. reduced form models). In this way the economic theory loses its role of guide for the hypothesis that should be tested through econometric analysis.

Somehow in line with this atheoretical but quantitative spirit, many recent empirical works make use of simple analysis of probability distributions of key microeconomic variables, simply aimed at describing as many stylized facts as possible by means of quantitative tools. What is usually found is the high degree of heterogeneity among micro-units. This fact prevents the use of the Law of Large Numbers to aggregate microeconomic variables and rules out the hypothesis of the representative agent.

It is not by chance that I chose factor models as the topic of my dissertation. These models indeed provide us with the tools to conduct econometrics research in the spirit that I outlined in this chapter. They allow to deal with large datasets that play the role of substitutes of random experiments outcomes. They are completely atheoretical and indeed were firstly introduced in economics by Sims in 1977. And finally they were proposed by Mario Forni and Marco Lippi in 1997 as tools for aggregating heterogeneous microeconomic variables. It is a chance that these models were developed in the 1930s by Karl Pearson and Charles Spearman in the field of psychology which in the 1970s gave many contributes to decision-making theory. The study of the human brain is to economics as the study of quarks is to physics.

---

<sup>3</sup>Other sciences may suffer the same problem, but either have many more data (as astronomy) or do not have any pressure from policy implication that biases the results (as cosmology).

## Bibliography

- J. S. Bell. *Speakable and unspeakable in quantum mechanics*. Cambridge University Press, 1987.
- A. L. Bowley and A.R. Burnett-Hurst. *Livelihood and Poverty: a study in the economic conditions of working-class households in Northampton, Warrington, Stanley and Reading*. 1915.
- E. Engel. Die Produktions und Consumptions Verhältnisse des Königreichs Sachsen. *Zeitschrift des Statistischen Bureaus des Koniglich Sachsischen Ministerium des Innern*, 22, 1857. Transl: The relations of production and consumption in the Kingdom of Saxony.
- R. P. Feynman and A. R. Hibbs. *Quantum mechanics and path integrals*. McGraw-Hill, 1965.
- M. Forni and M. Lippi. *Aggregation and microfoundations of dynamic macroeconomics*. Oxford University Press, 1997.
- T. Haavelmo. The probability approach in econometrics. *Econometrica*, 12 Supplement, 1944.
- D. Kahneman and A. Tversky. Prospect theory: an analysis of decisions under risk. *Econometrica*, 47:313-327, 1979.
- J. M. Keynes. *A treatise on probability*. Macmillan London, 1921.
- E. Majorana. Il valore delle leggi statistiche nella fisica e nelle scienze sociali. *Scientia*, 36:58-66, 1942. Transl: The value of statistical laws in physics and social sciences.
- A. Marshall. *Principles of economics*. Macmillan London, 1890.
- J. C. Maxwell. On the dynamical theory of gases. *Philosophical Transactions of the Royal Society of London*, 157:49-88, 1867.
- C. Menger. *Grundsätze der Volkswirtschaftslehre*. 1871. Transl: Principles of Economics.
- R. Nelson and S. Winter. *An evolutionary theory of economic change*. Harvard University Press, 1982.
- A. Roncaglia. *La ricchezza delle idee. Storia del pensiero economico*. Editori Laterza, 2003.
- J. Schumpeter. *Wesen und Hauptinhalt der theoretischen Nationalökonomie*, 1908. Transl: The nature and essence of theoretical economics.
- H. A. Simon. *Models of thought*. Yale University Press, 1979.
- C. A. Sims, Macroeconomics and methodology. *Journal of Economic Perspectives*, 10:105-120, 1996.
- J. Tinbergen. *Statistical Testing of Business-cycle Theories: Business Cycles in the United States of America 1919-1932*. League of Nations Publications Genève, 1939.