MEMORANDUM

No 07/2010

The “Meteorological” and the “Engineering” Type of Econometric Inference: a 1943 Exchange between Trygve Haavelmo and Jakob Marschak

Olav Bjerkholt

ISSN: 0809-8786

Department of Economics
University of Oslo

This series is published by the
University of Oslo
Department of Economics

P. O.Box 1095 Blindern
N-0317 OSLO Norway
Telephone: + 47 22855127
Fax: + 47 22855035
Internet: http://www.oekonomi.uio.no
e-mail: econdep@econ.uio.no

In co-operation with
The Frisch Centre for Economic Research

Gaustadalleén 21
N-0371 OSLO Norway
Telephone: +47 22 95 88 20
Fax: +47 22 95 88 25
Internet: http://www.frisch.uio.no
e-mail: frisch@frisch.uio.no

Last 10 Memoranda

No 06/10	Dag Kolsrud and Ragnar Nymoen *Macroeconomic Stability or Cycles? The Role of the Wage-price Spiral*
No 05/10	Olav Bjerkholt and Duo Qin *Teaching Economics as a Science: The 1930 Yale Lectures of Ragnar Frisch*
No 04/10	Michael Hoel *Climate Change and Carbon Tax Expectations*
No 03/10	Geir B. Asheim *Comparing the Welfare of Growing Economies*
No 02/10	Rolf Golombek, Mads Greaker and Michael Hoel *Climate Policy without Commitment*
No 01/10	Geir B. Asheim, Max Voorneveld and Jörgen W. Weibull *Epistemically Stable Strategy Sets*
No 27/09	Erik Biørn *Capital Decay and Tax Distortions: How to Abandon Exponential Decay and Benefit from It*
No 26/09	Erik Biørn *Modelling Addictions in Life-Cycle Models:Revisiting the Treatment of Latent Stocks and Other Unobservables*
No 25/09	Erik Hernæs and Zhiyang Jia *Labour Supply Response of a Retirement Earnings Test Reform*
No 24/09	Tarjei Havnes and Magne Mogstad *Money for Nothing? Universal Child Care and Maternal Employment*

Previous issues of the memo-series are available in a PDF® format at:
http://www.oekonomi.uio.no/memo/index.html

The “meteorological” and the “engineering” type of econometric inference: a 1943 exchange between Trygve Haavelmo and Jakob Marschak

by

Olav Bjerkholt

Department of Economics, University of Oslo / Statistics Norway

Abstract

The article presents an exchange of letters between Jakob Marschak and Trygve Haavelmo in May-July 1943. Marschak had from the beginning of 1943 become the research director of Cowles Commission at the University of Chicago. Trygve Haavelmo, who at the time worked for the Norwegian Shipping and Trade Mission in New York, had just published the article on the statistical implications of simultaneous equations, which would become his most quoted work. The content and the implications of the article was at the centre of the letter exchange. The introduction provides some background for the exchange.

JEL classification: B23
Key words: history of econometrics

Introduction

Econometrics, conceived as the idea of quantifying economic relationships, has roots far back in history. The history of econometrics literature, written largely in the 1980s and 1990s, has little controversy about what constituted the scientization of econometrics. Many forerunners have been given credit for their contributions to the development of econometrics, but the literature seems to be fairly unanimous about the fundamental importance of Trygve Haavelmo's contribution through his *Probability Approach in Econometrics* (Haavelmo, 1944), which according to Spanos (1989) is “commonly acknowledged as having founded modern econometrics.” Equally, or even more, frequently quoted is another contribution by Haavelmo, his Econometrica article on the statistical implications of systems of simultaneous equations (Haavelmo, 1943).[1]

The foundation of scientific disciplines springs from scientific breakthroughs and innovative ideas, sometimes rooting the new discipline in established theory and/or methodology in another field, as indeed econometrics may be viewed as rooted in probability theory and statistics. The new discipline may benefit from an institutional setting. The literature seems equally unanimous about what that setting was in the case of econometrics, namely the Cowles Commission at the University of Chicago under the leadership of Jakob Marschak from 1943.

The Cowles Commission can hardly be said to have been a very important research centre in economics prior to Marschak's directorship. It had been established by Alfred Cowles in Colorado Springs in 1932, as a foundation to support the activities of the then newly established Econometric Society. Cowles became Secretary and Treasurer of the Society and provided most valuable financial support by covering for several years the deficits of the Society's journal Econometrica. The Commission had a small scientific staff, had published 5-6 Monographs on different topics and for some years arranged research conferences at which the participants spent an entire month together in Colorado Springs. It also served as office for the Managing Editor of Econometrica. In 1939 the Cowles Commission moved from Colorado Springs to the University of Chicago with offices prominently located in the new Social Science Building.[2]

It was Marschak with the excellent team he managed to put together and the key programmatic role played by Haavelmo's path breaking contribution which gave the Cowles Commission its pivoting role in the history of econometrics, resulting in a massive thrust forward in econometric methodology.[3] Coinciding with Marschak taking over as Research Director at the Cowles Commission Haavelmo (1943) appeared in Econometrica in the January issue of 1943. Soon after Marschak had established himself in Chicago he studied Haavelmo’s paper. Marschak was certainly no novice in this field, nor was he unacquainted with Haavelmo and his work.

---

[1] Spanos (1989, p.405). Morgan (1990) likewise calls the publication of Haavelmo (1944) a ‘probabilistic revolution’ in econometrics (p.229), Qin (1993) uses similar words and quotes contemporary authors calling it the ‘logical foundations of econometrics’, ‘the manifesto’ of econometrics etc. (pp.20-21). Christ (1994) calls Haavelmo (1943,1944) jointly as ‘pioneering contributions’, and the latter one as setting forth the approach followed by the Cowles group.

[2] About Cowles Commission in the early years, see Cowles Commission (1952).

[3] Christ (1994) puts Tjalling Koopmans on an equal footing with Haavelmo as chief originator of Cowles’ theoretical econometric work.

The letters which make up the main content of this article, are letters exchanged in the period May-July 1943 between Marschak and Haavelmo. The content is entirely devoted to the clarification of the content, meaning and implications of Haavelmo (1943). Marschak's immediate motivation, as apparent from the letters, was for applications in demand analysis which had been his main research interest for some years. He had in fact a year and a half earlier made an effort to engage Haavelmo to work with him on the project (see below).

The Marschak-Haavelmo 1943 exchange can be read as Marschak interpreting, reformulating and proposing extensions to the results of Haavelmo (1943), while Haavelmo calmly explains, corrects when necessary Marschak’s unwarranted jumps to conclusions, and provides missing proofs. Marschak’s immediate motivation for clarifying the implications of Haavelmo (1943) was with regard to the research project already alluded to. This was a project he had brought with him from his previous position at the New School for Social Research in New York. Thus over and beyond this project Marschak’s main responsibility was to direct research at Cowles Commission, which at the outset required a research program. Cowles Commission was a rather small research facility, but Marschak was probably well aware that further expansion was within reach. The relocation of Cowles Commission in 1939 from Colorado Springs, Colorado to the University of Chicago gave by itself new possibilities for the Cowles Commission research, in terms of interacting with the Department of Economics and also in other ways benefit from the resources of the University of Chicago.

The Marschak-Haavelmo nexus may thus be seen to be of particular importance for Marschak’s decision to promote econometric methods in the research program for the Cowles Commission. Marschak had had three predecessors as research directors at Cowles Commission, but none of them had succeeded in making much impact on the profession of the research in the Commission.

World War II caused redirection if not upheavals for many research institutions. Staff left to work for war agencies and research efforts were oriented towards war needs or post-war problems. To some extent this was the case also for the Cowles Commission. One may reasonably surmise that an overall concern for the Commission and its new research director would be adapt its research program to give maximum benefit for post-war needs.

Jakob Marschak had hardly come very far in hammering out a research program for the Commission by the time of the letter exchange and perhaps not until the end of 1943. After further recruitment in 1944, including Tjalling Koopmans, Marschak prepared a research conference at Cowles Commission for the first week of January 1945, a conference which eventually resulted in the famous Cowles Commission Monograph no. 10 volume. Haavelmo’s work came, as is well known, to play an instrumental role in this research program.

The letter exchange of May-July 1943 can reasonably be considered as the first discussion of Haavelmo's results at the Cowles Commission under Marschak.[4] Haavelmo had at the time no formal association with Cowles Commission, but an indirect outcome of the letter exchange was that Marschak proposed to the board of the Commission that Haavelmo was appointed as a research associate, although there was no precedence for absentee appointments.[5] Marschak, indeed, used the discussion he had going with Haavelmo as an argument at the Commission

[4] Although it is not apparent from Marschak’s letters that others at the Cowles Commission read the exchange, it seems likely that Leo Hurwicz would have been kept fully informed by Marschak and perhaps others as well.

[5] The appointment did not involve any salary and was as Marschak put it “a purely moral tie”, Marschak/Haavelmo, 19 July 1943.

for appointing Haavelmo. It may further be viewed as a consequence of the Marschak-Haavelmo exchange that Haavelmo's major opus was eventually published as an annex to Econometrica in 1944. Furthermore, the clarification resulting from the exchange of letters may also seem to have scientific consequences as it had an impact on the demand project that Marschak worked on and which drove most of the Marschak's queries in the exchange of letters, as well as on a slightly later production study (Marschak and Andrews 1944). These projects by Marschak may well have been the first attempt to apply Haavelmo's results.[6]

The prior connection between Marschak and Haavelmo, originating in Europe, may reasonably be said to be rooted in their respective relations to Ragnar Frisch, and is briefly set out below. Marschak (born 1898) was by the time of the foundation of the Econometric Society in 1930 an experienced and quantitatively oriented economist. At the time of the first European meeting of the Econometric Society in Lausanne in September 1931 he had just become *Privatdozent* at Heidelberg University. The detailed report from the meeting showed Marschak almost measuring up to Frisch in the number of interventions at the meeting.[7] Frisch and Marschak must have known about each other also before 1931, even if they had not met, but certainly from that time they had excellent relations and great respect for each other. Marschak, like several others of the younger generation who took part in the activities of the Econometric Society took great interest in some of Frisch's ideas in econometrics.

The much younger Haavelmo (born 1911) graduated in economics at Frisch's Institute in 1933 and became his research assistant from that time. He advanced to become a close collaborator of Frisch and took central part in all his major projects in the 1930s. Marschak continued to participate in the Econometric Society's European meetings and inter alia wrote the report from the important third meeting in Leyden in 1933, at which Frisch presented internationally for the first time his propagation-impulse explanation of business cycles and at which Marschak's own contribution was “theoretical problems suggested by Roosevelt's policy.”[8]

By the time of the 1933 meeting Marschak's career in Germany was over. He fled Heidelberg and Germany for Vienna in March 1933, to decide where to go next. High up on his list was Frisch's Institute in Oslo, and in an intensive letter exchange with Frisch in March 1933 he explored the possibilities. Frisch, naturally, was enthusiastic about the idea, but failed in persuading Rockefeller Foundation to assist. Instead Marschak travelled via Spain and the Netherlands to England where he became head of a new Institute of Statistics at Oxford University from 1935. In 1936 Frisch brought Haavelmo with him for the first time to the annual European Econometric Society meeting, which that year took place in Oxford. Haavelmo presented a paper written within Frisch's macrodynamic and confluence analytic paradigms. In the discussion of Haavelmo's paper Marschak commented upon Frisch's structural modelling approach, as applied by Haavelmo, and thus gave a Frisch a chance to elaborate upon the meaning of the concept of ‘autonomy’, although this term had not yet appeared in print.[9] Haavelmo took the concept over from Frisch and gave it prominence as a key concept in Haavelmo (1941, 1944). Although the term is not mentioned explicitly in the

---

[6] Possibly in competition with an early application by Lawrence Klein, who at the time was Ph.D. student at M.I.T. and had arranged a seminar in March 1943 for Haavelmo to present his general ideas including the ideas set out in Haavelmo (1943). See Klein/Haavelmo, 6 May 1943 and Klein (1991).

[7] *Econometrica*, 1, 73-86.

[8] *Econometrica*, 2, 187-203.

[9] *Econometrica*, 5, 373-374. Haavelmo's paper was eventually published as Haavelmo (1938). The term ‘autonomy’ appeared in Frisch (1938), an unpublished paper.

1943 letter exchange, neither in Haavelmo (1943), it is nevertheless present and motivates the discussion of the use of least squares method versus the maximum likelihood method.

In the spring of 1938 Haavelmo, still Frisch's assistant, visited on a six months study tour Berlin, then Tinbergen's group in Geneva, followed by Paris, and, finally, Marschak's Institute in Oxford.[10] One year later both Marschak and Haavelmo were visiting USA, Marschak had arrived already at the end of 1938, Haavelmo in June 1939. They met again at the Cowles Commission Research Conference in Colorado Springs in July 1939, at which also Abraham Wald, Gerhard Tintner and others interested in econometrics were present.[11] After the conference Haavelmo, Marschak, Wald and a couple of others remained for an improvised post-conference colloquium.

Haavelmo's study purpose in USA was statistical testing of structural relations in economic theory. Testing economic theory, using the recently developed Neyman-Pearson methodology, was the cornerstone of his research idea. Soon after he had arrived in USA he found it necessary to go much deeper into the role of probability in economic theory. His planned length of stay was one year and a half, he had no wish to stay any longer. The German occupation of Norway from April 1940 made, however, a return to Norway politically impossible. Marschak on the other hand decided to remain in USA after the outbreak of the World War II and was offered a position at the New School for Social Research in New York.

After the Colorado Springs conference Haavelmo spent time with Jerzy Neyman at Berkeley and visited the Cowles Commission in Chicago before he came to New York in early 1940 to work with Abraham Wald. During the spring of 1940 Haavelmo and Marschak also had close contact, Marschak was working on a contribution to a memorial volume for Henry Schultz which was not published until 1942, but Marschak dates his contribution unusually precisely as being written in March 1940.[12] Marschak's contribution has a number of suggestive ideas and is by Qin (1993, pp.102-104) considered as an 'insightful breakthrough' in identification theory. Marschak (1942) had in fact a discussion of the very same two-equation example as found at the outset of Haavelmo (1943).[13] Morgan (1990, p.216) considered the ideas in Marschak (1942) as based on a development of suggestions in Haavelmo (1938), the published version of his presentation at the Oxford meeting in 1936, while Qin (1993) treated Marschak's paper as a more original contribution taking place concurrently with Haavelmo's approach to identification. Both seem to have overlooked the somewhat unusual acknowledgement Marschak (1942) gave to Haavelmo. While acknowledging benefits from discussions with Lange, Mosak and Wald, Marschak stated unequivocally that the article "could not have been written without the stimulating influence of talks with T. Haavelmo." Hence, we may conclude that Marschak's article drew on discussions with Haavelmo in March 1940, but quite likely at Colorado Springs in the previous year, as well.

Haavelmo had by that time done a fair amount of work on the treatise that eventually became Haavelmo (1944). The use of the simple two-equation model for discussion of identification appearing both in Marschak (1942) and in Haavelmo (1943), can however, be traced to Frisch (1933), which Haavelmo knew well from his early years with Frisch, but was well-known

---

[10] See Bjerkholt (2005).

[11] See Bjerkholt (2007).

[12] See Marschak (1942), p.135.

[13] See equations (1.1)-(1.2) in the excerpts form Haavelmo (1943) below.

also to Marschak who had intervened, almost like a conciliator, to calm the 'acrimonious debate' between Frisch and Leontief in 1933-34.[14]

Also in 1940 Haavelmo and Marschak took part in the Cowles Commission Research Conference in July, the last such event under the shadow of Pike's Peak in Colorado Springs. Again were they joined by Wald and Tintner, but also by new participants, such as Wassily Leontief and Paul Samuelson. After the conference Haavelmo moved to Harvard and halfway into the following term he had completed by April 1941 a ms. on "the theory and measurement of economic relations", his study propose for the visit to USA. He made some final revisions to the ms. during the summer, which he spent hiking and discussing with Abraham Wald in Maine. He had the ms. mimeographed and distributed in a limited number of copies in September as Haavelmo (1941). By then he had completed his mission, but had no place to go. He had the rest of the year left of his Rockefeller Fellowship, so he retained student status and remained at Harvard University.

Marschak, knowing that Haavelmo's fellowship was about to expire, arranged for a position on the faculty of the New School for Social Research (contingent on support from Rockefeller Foundation). He offered also participation in the Rockefeller supported econometric research project on demand studies that he directed. Marschak wrote a note to the Rockefeller Foundation to ask for additional financial support to cover Haavelmo's participation characterizing as follows:

> … he is one of the very few men, who understand the distinction between two types of econometric problems: prediction of uncontrollable events, and estimation of effects of a given government policy – the "meteorological" and the "engineering" type of econometric inference. Instead of applying blindly traditional statistical techniques we have to adapt our techniques to the special type of problem in hand. … In particular, it is our common intention to check the results of our analysis of demand by scrutinizing the effects of actual policies … . Any practical recommendations of policy (…) must, if they are to be based on any measurements, be based on measurements appropriately made; the regression coefficients relevant to such problems of "social engineering" are often numerically quite different from the regression coefficients obtained for "meteorological" forecasts.[15]

The distinction, in Marschak's terminology, between the "meteorological" and "engineering" aspect of econometric inference, was a key idea in the search for better econometric methods. It certainly was an idea Marschak shared with Haavelmo, and it may well be that it was the contact with Haavelmo since 1939 that had helped Marschak reach that insight. Passive observations of changing "economic weather" might, however, using the most appropriate methods, reveal information about parameters of autonomous relations. These could be used for (social) "engineering".[16] The prospect for social engineering applications was surely an integral part of Marschak's interest in better econometric method. For both Marschak and

---

[14] See Morgan (1990, pp.186-187), Hendry and Morgan (1990, pp.38-40).

[15] J. Marschak: Note on the continuation of a research project in econometrics, Nov. 6, 1941. Typewritten. Haavelmo Archive.

[16] There is in Marschak's note to Rockefeller Foundation 1941 also a passage that may seem to foreshadow the simultaneity problem of Haavelmo (1943): "The interaction between the variables, even in the most simplified form of the "three Keynesian equations" makes it necessary to be very cautious in applying the techniques of statistical estimation and forecasting to any general 'macroeconomic model'." But this general warning about estimating a simultaneous model can surely be traced back to Frisch.

Haavelmo economic and statistical research had no meaning unless it was for the betterment of society, perhaps something that was so obvious that it was taken for granted.

Haavelmo accepted without hesitation Marschak's offer of teaching and research at the New School, but had barely started to work on Marschak's project when he was called upon in January 1942 by the Norwegian Government-in-exile to serve as a statistician at the Shipping and Trade Mission in New York. Thus during 1942 Haavelmo lived and worked in Manhattan and had practical opportunities for keeping in touch with Marschak at the New School. Marschak had in 1940 initiated a seminar in mathematical economics and econometrics. In 1942 it met regularly on a Saturday every second month. Haavelmo gave two of the seminars that year, in February and in December, naturally, speaking on the topics he had dealt with in his 1941 treatise.[17] It was during a brief hospitalization for appendectomy between these two seminars that he drafted Haavelmo (1943).

Towards the end of 1942 Marschak decided to accept an offer of a chair at the University of Chicago and at the same time become Research Director of the Cowles Commission. Exactly why and how this happened is somewhat unclear, but Oskar Lange had surely exerted an influence.[18] Thus Marschak moved to Chicago from 1943, and there he came to work quite closely with Lange who also had a dual appointment. Lange was at the time serving as Acting Editor of Econometrica while Frisch was incommunicado in Norway (and imprisoned 1943-44).

Haavelmo had by the time of the letter exchange used his spare time over some months to revise the mimeographed Haavelmo (1941). He had surely been encouraged and pushed on by some of his friends to have the treatise properly published. It had after all only been distributed to a small number of persons and hardly been widely read. He completed the revision in June and showed the manuscript first to Oskar Lange. Marschak, learning this from Lange and Haavelmo, came up with the idea of publishing Haavelmo's treatise together with Mann and Wald (1943) as a Cowles Commission Monograph.[19] This was much to Haavelmo's liking, but turned out to be infeasible due to the wartime paper rationing. Instead Marschak, Oskar Lange and Dickson Leavens, the Managing Editor of Econometrica, arranged together for the treatise to be printed as an (unprecedented) annex to Econometrica, indeed a better way of publishing it.

Haavelmo (1943) instead of using Marschak's "meteorological" versus "engineering" terminology, spoke of the econometrician being in the situation of the astronomer unable to interfere with the course of events versus that of a planner having the power to change certain aspects of the economy. The econometrician as astronomer needed "equations of prediction",

[17] The titles of the two seminars were "The Nature and Logic of Econometric Inference" and "Problems of Estimation and Prediction in Economic Dynamics", see Bjerkholt (2007). One of those who attended the first of these two seminar commented upon the event later: "Many of us knew that an important turning-point had been reached, and Marschak saw the need for effective leadership." Arrow (1978, p.71).

[18] Lange had been appointed assistant professor in 1938. After the death of Henry Schultz in 1938 Lange was the only mathematical economist and econometrician in the department. After Paul Douglas had left to fight the war there was not much quantitative interest left in the department and also for that reason Marschak was a good choice from Lange's point of view, an earlier attempt by him to get the department to hire Abba Lerner, had failed, see Reder (1982). The board of Cowles Commission had considered Marschak as a suitable candidate for the research director position when Charles Roos resigned in 1937 and offered it to him. Marschak declined however, he was still living in England at that time.

[19] Interspersed with the letters on econometric issues Marschak and Haavelmo exchanged letters about the more mundane issue of how to publish what eventually became Haavelmo (1944).

while the econometrician as planner needed "equations of theory". This distinction was indeed the topic of the last section of Haavelmo (1943).[20]

Haavelmo may at this time, as his terminology suggests, very well have imagined himself as a planner in post-war Norway. Although his governmental work was not related to post-war reconstruction he was taking part in post-war planning on a more informal basis and was in frequent contact with economist friends in the Norwegian government-in-exile in London. Like many others he might have expected the war to end and Norway liberated sooner that turned out to be the case. To practice his insight in the use of econometrics relatively soon for social engineering, as part of an overall planning effort may well have seemed realistic for him at the time. He had no wish to stay in United States any longer than for the war to end. Haavelmo would surely have characterized himself as a socialist at the time, in a Norwegian context he adhered to the Labour Party, a social democratic party destined to become the dominant political force in post-war Norway. Some of his friends in the small econometric circuit may have held more leftish views. They may all have shared interest in putting econometrics to work for stabilization planning and welfare oriented policies.

The meteorologist-engineer or astronomer-planner distinction can also be traced – in other words – as a passing remark in Koopmans' first paper after returning to econometrics by being recruited by Marschak to Cowles Commission from his wartime shipping administration work: "It is true that the use to which we put estimates of the parameters of relations between economic variables is again one of prediction. But this is a different type of prediction, in which the effects of certain presumed acts of policy like price regulation, or instituting compensatory public works, or influencing saving habits, etc., are to be predicted. In such cases we are dealing with a type of prediction in which one or more of the relations found to govern the past are altered, and which is therefore not a straight forecast assuming continuation of all past relationships." (Koopmans, 1945).

The meteorologist-engineer distinction was close to the core of the new probability founded structural modeling approach of the Cowles Commission. Marschak elaborated upon it in the first 2-3 pp. of his Introduction to the famous Cowles Commission Monograph No. 10., again drawing on the meteorologist-engineer simile: "The economist's objectives are similar to those of an engineer but his data are like those of a meteorologist." Then he summarized main tenets of the Cowles Commission approach, later recognized as Haavelmo's contribution, but at the time curiously presented just as "forgotten" knowledge. Economic statisticians had "often forgotten" the role of simultaneous equations when trying to estimate a single stochastic relation, while economic theorist were "apt to forget" that observed economic variables are in general stochastic and thus economic hypotheses must be formulated as statistical ones in terms of probability distributions, Marschak (1950, pp.2-3). The "social engineering" objective as a motivation factor became less emphasized in the highly technical econometric development that followed.

It is in this context that the exchange of letters took place over the content of Haavelmo (1943). The content of Haavelmo (1943) was, as succinctly stated by Christ (1994, p.151), "a short clear demonstration, by means of simple examples, of why least squares yields biased and inconsistent estimators in simultaneous equations models, and how to get consistent estimators in special cases that we now recognize as just identified." The discussion between Marschak and Haavelmo is down-to-earth, very technical and perhaps not so attractive to read for other than specialists. The letters have a number of references to Haavelmo (1943). Hence,

---

[20] Prediction was also the topic of the last chapter of the Haavelmo (1944), which apart from that chapter was only a mildly reedited version of Haavelmo (1941).

all the formulae referred to in Haavelmo (1943) are quoted in a prefatory section of the correspondence extracts below. Haavelmo (1943)'s use of "special cases" implied that Haavelmo also had to refer Marschak at times to his *major opus*. It had already been rewritten and re-edited as it would appear in Haavelmo (1944). But as Haavelmo possessed the only copy he had to refer Marschak to the corresponding sections of Haavelmo (1941), to be found in the Cowles Commission's small library.

As an afterthought one might ask how important Marschak-Haavelmo nexus was for the probabilistic revolution. It was stated in the announcement from the Royal Swedish Academy of Sciences in connection with the Nobel Memorial Prize in Economics awarded Haavelmo in 1989, that Haavelmo (1941) "had a swift and path breaking influence on the development of econometrics"[21] This can hardly be said to be true unless it is taken to mean a swift and path breaking influence through its corroboration, application and promotion by the team gathered at Cowles Commission under Marschak's directorship. Haavelmo's major opus did not reach very widely on its own, it was after all "a long and rather technical paper" (Christ, 1994, p.151).[22] Haavelmo had in the revision of Haavelmo (1941) changed very little and hardly made the final version significantly easier to digest for the general reader. Even the much simpler and easier accessible Haavelmo (1943) did "not seem to have received attention commensurate with its importance for our subject", according to Koopmans (1945), published almost three years later.

The first reviewer of Haavelmo (1944), the most distinguished statistician at Harvard at the time, Edwin B. Wilson, apparently had no belief in a successful reception of the message in Haavelmo (1944). In a rather chilling review Wilson (1946) blasted Haavelmo's approach for being "extremely abstract and metaphysical", for relying on "hypothetical and abstract" illustrations rather than concrete applications, and for presenting within his treatise probability theory to such a high degree of abstraction that it pedagogically was a "suicidal" approach. Even the language, Haavelmo's "somewhat foreign" English, was an obstacle in Wilson's view. Wilson refused to accept one of Haavelmo's main tenet, rendered by Wilson as "the backward science of econometrics must be more critical with respect to its probabilistic hypotheses than other sciences need be." Curiously, Wilson was one of very few to whom Haavelmo acknowledged debt in the preface of Haavelmo (1944). Wilson may have had his time as a promoter and supporter of the new discipline of econometrics and was an outlier among the reviewers. But surely, it would have been a much slower 'probabilistic revolution' arising from Haavelmo (1944) without the vehicle of the Cowles Commission.

The "social engineering" about which Marschak and Haavelmo both nurtured high expectations turned out to be more difficult to effectuate than it may have seemed in 1943. Haavelmo stated later on various occasions that economic theory needed further development for the econometric methods to become fully applicable. But Haavelmo's attempts to apply the 'Probability Approach' after the war and his return to theory building is another story.

---

[21] See *Scandinavian Journal of Economics* (1990), pp. 11-15. The Royal Swedish Academy of Sciences also stated that Haavelmo (1941) had been defended as a doctoral thesis at Harvard, which, if it had been correct may have been helpful in getting his work distributed and known, but its relation to Harvard University was just that it had been mimeographed on the premises. (Haavelmo was conferred his doctoral degree by the University of Oslo in 1946 for Haavelmo (1944).)

[22] In the period 1941-1944 I have found only a couple of references in the literature to Haavelmo (1941), namely one by Hans Staehle, who was Haavelmo's friend and even co-author, and one by Lawrence Klein.

The Marschak-Haavelmo 1943 exchange

Marschak and Haavelmo exchanged four letters in the period from 10 May 1943 until 8 July 1943. Marschak's letters were nicely typewritten by a secretary at the Cowles Commission. To the first letter he added a handwritten *post scriptum*, to the fourth he attached a four-page typewritten memo. Marschak addressed his the letters to Haavelmo's office at the Norwegian Shipping and Trade Mission on Broad Street in New York. Haavelmo refrained from using the secretarial capacity of the Norwegian Government for non-duty purposes and answered in longhand, written, we may presume, at late hours in his office or possibly at night in his home on W. 119 Street. From the letter dates it seems that Haavelmo answered each of Marschak's letters on the very same day he received it.

Marschak's need to fully understand the respective merits of maximum likelihood and least squares estimation made parts of the discussion repetitive. It has only been slightly edited and beyond notational changes and deletion of less relevant material. Some footnotes have been inserted for clarification, such as when the discussion went beyond Haavelmo (1943) and touched upon issues dealt with in Haavelmo (1941), the corresponding passages of Haavelmo (1944) have been indicated.

Haavelmo on the statistical implications of a system of simultaneous equations

Haavelmo (1943) contain some of the most frequently quoted equations in the econometrics literature, not least some of the equations quoted below. It began with the simplest possible system of simultaneous equations with random residuals.:

(1.1)-(1.2)
$$Y = aX + \varepsilon_1$$
$$X = bY + \varepsilon_2$$

The two random ε-variables were assumed to have zero means, variances equal to σ_1^2 and σ_2^2, and to be independent. Haavelmo showed the implication of (1.1)-(1.2) with regard to the expected value of *Y* for given *X*, to be:

(1.5)
$$E(Y \mid X) = \frac{\sigma_Y}{\sigma_X}\rho_{XY} X = \frac{b\sigma_1^2 + a\sigma_2^2}{b^2\sigma_1^2 + \sigma_2^2} X$$

Haavelmo (1943) then considered the equation system

(2.5)-(2.7)
$$u_t = \alpha r_t + \beta + x_t$$
$$v_t = \kappa(u_t - u_{t-1}) + y_t \qquad t = 1, \ldots, N$$
$$r_t = u_t + v_t$$

where u_t was consumption, v_t investment, r_t total income, and x_t and y_t random variables with zero means and variances equal to σ_x^2 for all the *x*'s and σ_y^2 for all the *y*'s.

For the system (2.5)-(2.7) Haavelmo showed that $E(u_t \mid r_t)$ was not $\alpha r_t + \beta$ as (2.5) might suggest. When the *x*'s and *y*'s were jointly normally distributed, u_t and r_t also were jointly normally distributed and $E(u_t \mid r_t)$ became a linear function $Ar_t + B$ with $A = \frac{\sigma_u}{\sigma_r}\rho_{ur}$ which could be worked out to be

(3.7)
$$A = \frac{(1+\kappa)\sigma_x^2 + \alpha\sigma_y^2}{(1+\kappa)^2\sigma_x^2 + \sigma_y^2}$$

Haavelmo further showed that the maximum likelihood estimates of $\alpha, \beta, \kappa, \sigma_x^2, \sigma_y^2$ in this model followed from the first-order conditions:

$$
\begin{aligned}
&\Sigma_1^N \{[u_t - \hat{\alpha} r_t - \hat{\beta}][r_t + (1+\hat{\kappa})(\hat{\beta} - u_t)]\} = 0 \\
&\Sigma_1^N \{[v_t - \hat{\kappa} u_t + \hat{\kappa} u_{t-1}][(1-\hat{\alpha})(u_t - u_{t-1}) - \hat{\alpha} v_t]\} = 0 \\
\text{(3.10)-(3.14)}\quad &\Sigma_1^N u_t - \hat{\alpha}\Sigma_1^N r_t - N\hat{\beta} = 0 \\
&\Sigma_1^N [u_t - \hat{\alpha} r_t - \hat{\beta}]^2 = N\hat{\sigma}_x^2 \\
&\Sigma_1^N [v_t - \hat{\kappa}(u_t - u_{t-1})]^2 = N\hat{\sigma}_y^2
\end{aligned}
$$

The last section of Haavelmo (1943) considered prediction problems using the same model, exemplifying by four prediction problems with corresponding prediction formulae.

To predict u_t from past observations:

$$\text{(4.1)}\qquad E(u_t \mid u_{t-1}) = -\frac{\alpha\kappa}{1-(1+\kappa)\alpha} u_{t-1} + \frac{\beta}{1-(1+\kappa)\alpha}$$

To predict v_t from past observations:

$$\text{(4.2)}\qquad E(v_t \mid u_{t-1}) = -\frac{(1-\alpha)\kappa}{1-(1+\kappa)\alpha} u_{t-1} + \frac{\kappa\beta}{1-(1+\kappa)\alpha}$$

To predict u_t when r_t and past observations were given:

$$\text{(4.3)}\qquad E(u_t \mid r_t, u_{t-1}) = \frac{(1+\kappa)\sigma_x^2 + \alpha\sigma_y^2}{(1+\kappa)^2\sigma_x^2 + \sigma_y^2} r_t + \frac{(1+\kappa)\kappa\sigma_x^2}{(1+\kappa)^2\sigma_x^2 + \sigma_y^2} u_{t-1} + \frac{\beta\sigma_y^2}{(1+\kappa)^2\sigma_x^2 + \sigma_y^2}$$

To predict v_t when u_t and past observations were given:

$$\text{(4.4)}\qquad E(v_t \mid u_t, u_{t-1}) = \frac{\kappa\sigma_x^2 + \alpha(1-\alpha)\sigma_y^2}{\sigma_x^2 + \alpha^2\sigma_y^2} u_t - \frac{\kappa\sigma_x^2}{\sigma_x^2 + \alpha^2\sigma_y^2} u_{t-1} - \frac{\alpha\beta\sigma_y^2}{\sigma_x^2 + \alpha^2\sigma_y^2}$$

Again using the same example Haavelmo illuminated the significance of the theoretical equations obtained by omitting the error terms in (2-5)-(2.6), by assuming that in a government planning context public spending g_t would become an additional term in (2.7), say,

$$\text{(2.7')}\qquad r_t = u_t + v_t + g_t$$

If, as a government policy decision, r_t were kept constant through varying g_t accordingly, (2.7') would not impose any restriction on u_t and v_t. From (2.5)-(2.6) would under this assumption then follow

$$
\begin{aligned}
\text{(4.5)-(4.6)}\qquad &E(u_t \mid r_t) = \alpha r_t + \beta \\
&E(v_t \mid u_t - u_{t-1}) = \kappa(u_t - u_{t-1})
\end{aligned}
$$

JM to TH: Chicago, May 10, 1943

Dear Haavelmo:

As you know, I am greatly interested in your, Wald's and Mann's work of which a part was contained in your recent Econometrica article and another part will be soon, I hope, brought into systematic form by Wald and Mann.[23] Have you been in touch with them recently and how is the work progressing? How was Mann's paper on the subject, as presented at the Seminar last month?[24]

In the meantime, I am working my way through your article—as you have seen from Leavens' letter on what Hurwicz and I thought was an erratum. Thank you very much for clearing up this point.[25]

I have today a few other questions on the subject.

<u>Question 1.</u> Referring to your equations (4.5)-(4.6) on page 12, Econometrica, January, 1933[!]. This is the case where the maximum likelihood estimates of the parameters are applied directly to answer a practical question. As you know, however, it is not always possible to estimate each parameter. For example, the system (analogous to your equations (1.1), (1.2) but rewritten for convenience so as to be symmetrical in the parameters)

$$\text{(I)} \quad \begin{aligned} Y_t &= aX_t + \varepsilon'_t \\ Y_t &= bX_t + \varepsilon''_t \end{aligned} \qquad t=1,...N$$

yields, when the likelihood of a is maximized, the same relationship as when the likelihood of b is maximized, viz.,

$$\text{(II)} \quad \Sigma Y^2 - (\hat{a} + \hat{b})\Sigma XY + \hat{a}\hat{b}\Sigma X^2 = 0$$

On the other hand, consider the system (where c is known and where Z_t are observations of a third variable which is "exogenous")

$$\text{(III)} \quad \begin{aligned} Y_t &= aX_t + cZ_t + \varepsilon'_t \\ Y_t &= bX_t + \varepsilon''_t \end{aligned}$$

This system does not lead to such indeterminacy as the preceding one, but another difficulty arises: $\hat{a}$ and $\hat{b}$ turn out to be the two roots of a quadratic equation and it is not possible without further information (e.g. a priori knowledge of the signs) to say which is which.

My question is: have you investigated under what conditions indeterminacy arises? The condition may be analogous to conditions of "multicollinearity", at least when the variances are not given. Do the five equations (3.10)-(3.14) yield determinate solutions? If so, are they unique? (I notice that (3.10) is quadratic in $\hat{\beta}$).

<u>Question 2.</u> This question is perhaps even more fundamental to my understanding of the subject. Equations (4.1)-(4.4) and, for a simpler case, equation (1.5) give conditioned expected values. I suppose that you recommend that, in practice, the maximum likelihood estimates of the parameters - as obtained from (3.10)-(3.14) - be inserted in the equations instead of the parameters themselves, thus giving the desired estimates of the conditioned expected values. Have you performed the operations for your examples? Would you then not

---

[23] The references are to Haavelmo (1943) and Mann and Wald (1943), the latter had not yet appeared.

[24] Henry B. Mann presented Mann and Wald (1943), not yet published, to the Marschak seminar in New York in the spring of 1943, after Marschak had moved to Chicago.

[25] Marschak and Leonid Hurwicz at their first reading of Haavelmo (1943) found that formula (3.16), the first-order condition for maximum likelihood estimation in the most general case Haavelmo considered, had to be wrong. Marschak instructed Leavens to get from Haavelmo an erratum notice for the next issue of Econometrica. The formula was correct, but the episode arose perhaps from Haavelmo's habit of using few words to accompany his formulae. Formulae (3.10)-(3.14) are special cases of (3.16). See letters Leavens/Haavelmo, 20 April 1943; Haavelmo/Leavens, 24 April 1943.

obtain, as the coefficient of, say, u_{t-1} in (4.1) simply the least squares partial regression coefficient of u_t on u_{t-1}? Does not A in (3.7) become simply the least squares partial regression of u_t on r_t, as soon as the maximum likelihood estimates are inserted for each of the respective parameters? In the example (I) above this is definitely the case, owing to the relationship (II) between the maximum likelihood estimates. Since you have not stated this directly in any place in the article, I still have doubts whether I understand the procedure you recommend for "predictions" as on page 11.[26]

On the other hand, - to return to question 1 - I understand for "policies", the maximum likelihood, and not the least squares, estimates must be used as coefficients; and the trouble arising there is merely that of indeterminacy in certain cases, and of symmetrical roots in others. A systematic discussion of such cases would be very useful.

…

I suppose you are very busy, and I would be happy if you would answer these questions separately, one at a time, rather than postponing the whole answer. I am anxious to apply your results to the practical problem at hand: the work we did with Garvy last year is now being brought into shape, and I should like to make as much use of your theory as possible.[27] At least for the "lag-less" setup—which is analogous to the system (III) above, Z being income and c being the income-elasticity derived from budgets while X and Y are quantity and deflated price—this should not be difficult, but I am anxious to be on absolutely sure theoretical ground before starting any computations. Question 2 is more urgent one.

I hope that you are well and that we shall see each other before too long!

Sincerely,
J. Marschak

[handwritten]P.S. On page 8, para 2 you state that a function of maximum likelihood estimates is equal to the max. likelihood estimate of the function. Call this "Theorem 1". It is further known that, in the case of joint normal distribution, the max. likelihood estimate of the regression coefficient is the least squares estimate. Call this "Theorem 2". Write your equation (1.5) in the form

$$E(Y \mid X) = AX = F(\sigma_1, \sigma_2, a, b) \cdot X$$

indicating thus that A is a function of the population parameters. Then, using ^ to denote maximum likelihood estimates,

$$\hat{E}(Y \mid X) = \hat{A} \cdot X$$

= (because of Theorem 1) $F(\hat{\sigma}_1, \hat{\sigma}_2, \hat{a}, \hat{b}) \cdot X$.

But, because of Theorem 2,

$$\hat{E}(Y \mid X) = X \cdot \Sigma X_t Y_t / \Sigma X_t^2$$

Hence $\quad F(\hat{\sigma}_1, \hat{\sigma}_2, \hat{a}, \hat{b}) = \Sigma X_t Y_t / \Sigma X_t^2$

Specifically (as an illustration) : Assuming in the system (I) (p.1 of this letter) that $a < 0, b > 0$; the 2*N*-variate normal distribution of X_t, Y_t (t=1,…N) becomes

$p = (b-a)^N (2\pi\sigma_1^2\sigma_2^2)^{-N/2} \exp(-\frac{1}{2}\Sigma\frac{\varepsilon_1^2}{\sigma_1^2} - \frac{1}{2}\Sigma\frac{\varepsilon_2^2}{\sigma_2^2})$; this gives $N\hat{\sigma}_2{}^2 = \Sigma(Y - bX)^2$, and analogous for $N\hat{\sigma}_1{}^2$; while the differentiation of p with respect to b gives, when equating to

---

[26] I.e. the derived prediction formulae (4.1)-(4.4).

[27] Marschak worked with George Garvy, an economist of Russian extraction, on the demand studies project he had invited Haavelmo to join.

zero and inserting the values of $\hat{\sigma}_2$ just given, the equation (II); (the same equation is obtained when differentiation is made with respect to a). On the under hand[!],
$E(Y \mid X) = X \cdot \rho_{XY} \frac{\sigma_X}{\sigma_Y} = X \cdot \frac{b\sigma_1^2 + a\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$; inserting $\hat{\sigma}_1, \hat{\sigma}_2, \hat{b}, \hat{a}$, and using (II) we obtain as an estimate of $\rho_{XY} \frac{\sigma_Y}{\sigma_X}$ simply the expression $\frac{\Sigma XY}{\Sigma X^2}$.

TH to JM: New York, May 13, 1943

Dear Marschak:

Thank you very much for your letter of May 10. I am certainly happy to see that there is some interest in my article. I have also had some talks recently with Wald and Mann, and I understand that their work on this subject is practically finished and will appear in Econometrica perhaps in the next issue. From Mann's lecture at our seminar I got the impression that they have "cleared up everything", practically speaking, at least for large samples, and that the maximum likelihood method seem to be a good method in all cases.

I shall try to answer your questions, as far as I am able to, hoping you will forgive me for saving time by writing in pencil.

Question 1

I agree with your remarks on the possibility of indeterminate coefficients. This is by itself a whole chapter, and a very important one, which I did not feel like touching upon in my short article. E.g. the 5 equations (3.10)-(3.14) will, I believe, determine all the parameters involved, apart, perhaps from a certain indeterminacy as to which is which of the parameters, what their signs are or some similar trouble, that has to be remedied by some extra information. But I have not actually investigated this. In [Haavelmo, 1941] (if you should not have your copy, there is one in the Commission library) I have given a general method for investigating such problems (in Part IV). Essentially, the rule is that in order to be able to estimate the parameters of a certain distribution the partial derivatives of the distribution function with respect to all the unknown parameters have to be linearly independent.

Question 2

I understand you to mean that for such "expected value equations" as e.g. (4.1)-(4.4) one might just as well fit these equations directly to the data by the least squares method, writing e.g. (4.1) as $u_t = A_1 u_{t-1} + A_0 + error$ and letting A_1 and A_0 be the regression coefficients obtained. That is correct, with certain reservations which I shall mention below.

My point was not that one should never use the least squares method, but that one should not apply it to equations where the expected value of the dependent variable is different from the linear form one actually wants to estimate. I shall try to make my point clear.

Consider the equations (2.5)-(2.6) (and (2.7)). From such a system one can express u_t and v_t as linear functions of lagged values of the same variable and present and lagged values of the other variable plus an error term (containing x_t and y_t). And this can be done in many different ways. E.g. the original system (2.5)-(2.6) is one of them. The equations from which (4.1)-(4.4) follow are other examples.

Among these equations consider those that have the following property, called (*):

> The expected value of the variable on the left side given the variables on the right side is the same linear function of the right hand side variables as the one obtained simply by omitting the error term appearing in that equation. The condition for this is,

essentially, that the error term be uncorrelated with the observable variables in the right hand side of the equation in question.

For these equations, fulfilling (*), the least squares method can be applied directly, in the sense that it will at least give consistent estimates of the coefficients in the equation. The equations from which (4.1)-(4.4) were derived fulfill (*). They might therefore be fitted directly, introducing some new symbols for the compound coefficients so that (4.1)-(4.4) become linear in these new coefficients.

The equations (2.5)-(2.6) do not fulfill (*). And that was my main point. $E(u_t \mid r_t)$ is not equal to $\alpha r_t + \beta$. Therefore, we cannot fit (2.5)-(2.6) by the least squares method if we want to estimate α, β and κ. True, there are relations, implied by (2.5)-(2.6), that have the same form as (2.5)-(2.6) and which fulfill (*), but their coefficients are not simply α, β and κ but some functions of these parameters and the σ's.

My first reason to prefer the maximum likelihood method to estimate α, β and κ directly is that then one also has the max. likelihood estimates of whatever functions of the parameters we might like to consider, e.g. those that appear as coefficients in (4.1)-(4.4), simply by substituting. One does not have to do a new fitting process for every new equation implied by (2.5)-(2.6) that one might like to consider.

My second reason for suggesting the use of the maximum likelihood estimates also for deriving the coefficients in (4.1)-(4.4) is this: Consider the 2 independent equations (4.1)-(4.2). From the point of view of a direct application of the least squares method there are here 4 coefficients to be estimated, while actually only three unknown parameters are involved, namely α, β and κ. The maximum likelihood method takes account of this, the least squares method would not. Therefore the insertion of the maximum likelihood estimates in (4.1)-(4.2) does not yield the same coefficients as the least squares method. Your remark about A in (3.7) is therefore probably not correct. Otherwise I have no principal objection to fitting (4.1)-(4.4) by the least squares method. I am not yet sure which method gives the best estimates. Both are consistent.

But whenever the least squares method is applied, (*) must be fulfilled, otherwise one obtains only biased estimates of the coefficients one believes one is estimating.

The main reason why I put up the complicated relations (4.1)-(4.4) instead of simple linear forms was that I wanted to emphasize that the equations to be used for prediction purpose are not the original equations with error terms omitted.

…

Well, I hope you will be able to read these notes and that I have at least hit some of the problems which you have raised. Thank you again for the interesting letter.

Cincerely[!] yours
Trygve Haavelmo

JM to TH: Chicago, June 4, 1943

Dear Haavelmo:

Following your answer to Question 1 of my letter of May 10, I shall study Part IV of [Haavelmo, 1941].[28]

As to Question 2: my handwritten postscript contained a proposition which was much more general than the propositions in your answer to that question. If I am right, the use of least squares estimates of the coefficients for purposes of prediction, is permissible not only in

---

[28] Haavelmo (1941, Part IV) is practically identical to Haavelmo (1944, Chapter V).

single cases you mention but also under more general conditions. I should like to repeat my statement: although it is confined to two variables, its extension would seem obvious.

Let the $2N$ observations $X_t, Y_t \quad (t = 1, ...N)$ be connected by $2N$ equations

$Y_t = b_i X_t + a_i + \varepsilon_{it} \quad (i = 1, 2)$,

where ε_{it} are values of the normally distributed random variables ε_i with mean 0 and variances σ_i^2. Then the observations can be regarded as forming a sample taken from a joint normal distribution of two variables, X and Y. The expected value of Y given X is a linear function of X, say,

(1) $E(Y \mid X) = AX + B$

where the coefficients A and B are functions of the parameters of the joint distribution of the $2N$ observed values $X_1, Y_1, ...X_N, Y_N$; say,

$A = F(a_1, a_2, b_1, b_2, \sigma_1^2, \sigma_2^2) \quad B = G(a_1, a_2, b_1, b_2, \sigma_1^2, \sigma_2^2)$

Denote the maximum likelihood estimates of the six parameters by ^ superscripts. Further, define A^* and B^* as follows:

$A^* \equiv F(\hat{a}_1, \hat{b}_1, \hat{\sigma}_1{}^2, ...) \quad B^* \equiv G(\hat{a}_1, \hat{b}_1, \hat{\sigma}_1{}^2, ...)$

Then I think the following proposition can be proved:

(2) $$A^* = \frac{\sum_1^N (X_t - \bar{X})(Y_t - \bar{Y})}{\sum_1^N (X_t - \bar{X})^2} \qquad B^* = \bar{Y} - A^* \bar{X}$$

where $\bar{X} = \frac{1}{N}\sum_1^N X_t$, $\bar{Y} = \frac{1}{N}\sum_1^N Y_t$

i.e. A^* and B^* are equal to the least square estimates of the coefficients A and B in the regression equation (1).

The proof is based on two theorems:

Theorem 1: "a function of maximum likelihood estimates is equal to the maximum likelihood of the function." (I take this from page 8, paragraph 2 of your article; but I should be grateful for the proof of this theorem, or a reference).

Theorem 2: "the maximum likelihood estimates of the regression coefficient and of the constant term are, in the case of joint normal distribution, the least squares estimates." This follows directly from the equation of the joint normal distribution.

From Theorem 1 it follows that A^* and B^* are maximum likelihood estimates of A and B. But from Theorem 2 it follows that the maximum likelihood estimates of A and B in (1) are the least square estimates. Hence A^* and B^* are equal to the least squares estimates as stated in (2).

I should be grateful for your reactions.

…

Cordially yours,
J. Marschak

TH to JM: New York, June 7, 1943

Dear Marschak:

Thank you for your letter of June 4. You say in your letter that you were discussing (in your previous letter) a more general proposition than the one I was considering in my answer, while in fact I was trying to explain that the general proposition you mentioned (and which you repeat in this last letter) is not true under all circumstances.

Perhaps there might be some misunderstanding regarding "Theorem 1" (about the invariance of the maximum likelihood estimates). I just mentioned this theorem in my article without stating it very explicitly, since I have heard about it and used it so often that I thought it would be in the textbooks. But that is perhaps not so. Below I shall, therefore, try to state this theorem explicitly.

Let $X_1, X_2, ..., X_n$ be n random variables having the joint distribution

(1) $$p(X_1, X_2, ..., X_n; \alpha_1, \alpha_2, ..., \alpha_k)$$

where the α's are unknown parameters. Let $x_1, x_2, ..., x_n$ be a sample of n observations, one for each corresponding X.

Consider the likelihood function

(2) $$p(x_1, x_2, ..., x_n; \alpha_1, \alpha_2, ..., \alpha_k)$$

Assumption 1: The maximum likelihood estimates $\hat{\alpha}_i$, $i = 1, 2, ..., k$, as derived from (2) exist and are unique within a certain domain, A, of the k-dimensional parameter space of the α's. (in certain cases A might be the whole parameter space.) Consider k new parameters, β_i, defined by a non-singular transformation

(3) $$\beta_i = \varphi_i(\alpha_1, \alpha_2, ..., \alpha_k) \quad i = 1, 2, ..., k$$

Assumption 2:

To the domain A in the α-space corresponds a certain domain B in the β-space such that there is a one-to-one correspondence between the parameter points α in A and the parameter points β in B, as defined by (3). This means that we can solve (3) for the α's, say,

(4) $$\alpha_j = \psi_j(\beta_1, \beta_2, ..., \beta_n) \quad j = 1, 2, ..., k.$$

Assumption 3:

The partial derivatives $\frac{\partial \psi_j}{\partial \beta_i}$, $i, j = 1, 2, ..., k$ exist and are continuous throughout B. (and similarly for $\frac{\partial \varphi_i}{\partial \alpha_j}$).

Now insert (4) into (2). We then get

(5) $$p(x_1, x_2, ..., x_n; \psi_1, \psi_2, ..., \psi_k)$$

To find the maximum likelihood estimates $\hat{\beta}$ of the β's we maximize (5) with respect to the β's, which leads to the following system of equations:

(6) $$\frac{\partial p}{\partial \psi_1}\frac{\partial \psi_1}{\partial \beta_i} + \frac{\partial p}{\partial \psi_2}\frac{\partial \psi_2}{\partial \beta_i} + ... + \frac{\partial p}{\partial \psi_k}\frac{\partial \psi_k}{\partial \beta_i} = 0 \quad i = 1, 2, ..., n$$

Let

(7) $$\hat{\beta}_i, \quad i = 1, 2, ..., n$$

be the solution of (6). Then we have

Theorem 1:

The solutions $\hat{\beta}_i$ are unique and satisfy

(8) $$\hat{\beta}_i = \varphi_i(\hat{\alpha}_1, \hat{\alpha}_2, ..., \hat{\alpha}_k), \quad i = 1, 2, ..., k$$

or

(9) $$\hat{\alpha}_i = \psi_i(\hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta}_k), \quad i = 1, 2, ..., k$$

That the $\hat{\alpha}$'s, as derived from (2), define $\hat{\beta}$'s, via (8), which satisfy (6) is obvious for these $\hat{\alpha}$'s, by definition make all the quantities $\frac{\partial p}{\partial \psi_i} = 0$. If some other $\hat{\beta}$'s should exist that would satisfy (6) some of the quantities $\frac{\partial p}{\partial \psi_i}$ would, for the corresponding $\hat{\alpha}$'s, be different from zero. But then (6) could not be satisfied unless the Jacobian $\left|\frac{\partial(\psi_1, \psi_2, ..., \psi_k)}{\partial(\beta_1, \beta_2, ..., \beta_k)}\right|$ would vanish, and this is excluded by the one-to-one correspondence between the $\hat{\alpha}$'s and the $\hat{\beta}$'s as given by (8) and (9).

Theorem 1 gives only a sufficient condition for the invariance of the maximum likelihood estimates. If the conditions imposed above are <u>not</u> fulfilled, your general statement may or may not be true. One can give examples of both.

Your example does not fulfill assumption 1 above, for the estimates of b_1, b_2, a_1, a_2 are not unique (there is indeterminacy). You have 6 parameters, while the normal distribution in your case is determined by 5 parameters, namely $\sigma_X^2, \sigma_Y^2, E(X), E(Y)$ and $E(XY)$. (There is trouble even if you know a_1 and a_2). In your case it is correct that A and B can be obtained either by the maximum likelihood method or by least squares. But that has nothing to do with theorem 1.

In my example (with the consumption and investment variables) there is more trouble, for there I have fewer original parameters than those occurring in the corresponding least squares equations (that is, I have more β's than α's, in the notation used above).

If I wanted to fit the two equations (4.1)-(4.2)

$$u_t = A_1 u_{t-1} + A_2 + error$$

$$v_t = B_1 u_{t-1} + B_2 + error$$

by the method of least squares, I would have to do this under a side condition since there is a functional relations between the parameters.

One must not take “theorem 1” to mean that one can increase or decrease the total number of parameters in a distribution or use singular transformations and still have the invariance property working. This is obvious, because how would one otherwise be able to use the theorem if there were not a one-to-one correspondence between the old and the new parameters, or if the new parameters were not independent? Then one could not insert the estimates into the transformation equations and solve, to see whether the theorem is true or false.

I realize now that it is rather difficult to have such discussion by mail. But I am willing to try again if you find the above unclear. I wish we could meet somewhere and have a really thorough discussion on these questions. I think I learn a lot by being forced to make things understandable.

Yours sincerely,
Trygve Haavelmo

JM to TH: Chicago, June 30, 1943

Dear Haavelmo:

On the last page (p.121) of [Haavelmo, 1941] I find a paragraph which again puzzles me.[29] It starts with the words, “If one wants to predict,” etc. and contains – as far as I see – the same statement as was made by me in my previous two letters. However, no general proof is given, and I wonder whether your last letter (of June 7) disproves the statement given on p. 121, or whether there is some difference between that statement and the one made by me: perhaps mine is too general?

---

[29] The paragraph referred to is practically identical to Haavelmo (1944), p.104, the second to last paragraph beginning “If we want to *predict* …”.

It would greatly help my studying your letter of June 7 if you could give me your reaction to this question. Thank you very much.

Sincerely.
J. Marschak

TH to JM: New York, July 2, 1943

Dear Marschak:

Thank you for your letter of June 30. The statement by me which you quote from [Haavelmo, 1941] (p.121) refers only to the particular example I have used (although it would be true for a whole class of similar examples). I once actually carried out the calculations in both ways for this example, and they check. The reasons why the two methods of calculating the equation in question give the same result in my example are the following:

1. For any multivariate normal distribution where all the parameters are unknown the expected value of one variable given the others can be estimated either by inserting the maximum likelihood estimates of the parameters into the expected value equation or by fitting the expected value equation (as a linear equation with unknown coefficients) by least squares.
2. My example fulfils the conditions of the invariance of the maximum likelihood estimates which I set forth in my letter of June 7.

To clarify the problem we are discussing I would suggest the following experiment: Take the simple case of n independent observations from a bi-variate normal distribution. It will in general involve 5 parameters: correlation coefficient, two variances and two means. Now assume that these five parameters are themselves certain functions of one single parameter. Calculate the maximum likelihood under these side constraints. Insert these estimates in the expected value of Y given X. Next use the least squares method directly, disregarding the side-conditions. Then the two results will in general not agree.

With best regards,
Trygve Haavelmo

JM to TH: Chicago, July 8, 1943

Dear Haavelmo:

Thank you very much for your letter of July 2nd. For myself this discussion is of a very great value indeed, and I only hope I am not imposing too much on you.

I don't know whether the way I am inducing you to find more final and systematic formulations by giving you examples to crack is the most fruitful one. But it so happens that this procedure arises out of the type of simple applied problems with which I am dealing now.

In an attempt to determine simultaneously demand and supply equations for a single commodity, I have applied your method in the attached note. It is, of course, only a first step. At the next step, the supply equation will include production cost as a further exogenous variable in the sense defined in Section 1 of the note. Further, a lag will be attached to the price variable (x_0) in the supply equation. As long as "small" commodities are considered, it is legitimate, I suppose, to regard certain variables as exogenous, and not to attempt to write down the full system. The joint distribution written down is that of endogenous variables only, with the exogenous ones as constants. This will be different when an attempt will be made later to formulate a fuller system with the variables "savings", "food", "rent", "other consumables", "investments", "industrial costs", etc.

Yet already at this stage I feel insecure, and this insecurity is concerned with the same two equations we have been discussing recently, viz.: A) the question of indeterminacy with respect to certain parameters, B) the question of conditions under which the predictions by inserting maximum likelihood estimates into the expression for $E(Y \mid X)$ gives results identical with prediction by least squares.

Question A). In section 4 of the enclosed note I define two cases: in the one, b_1 (income elasticity) is unknown, in the other it is known (from family budgets). I obtain maximum likelihood estimates in both cases: see Section 5 for the first case, and Section 8 for the second. But does this agree with the test described in Section 19 of your book?[30] In the second case, no indeterminacy arises, since in the demand equation (1.1b) a new variable can be introduced, $z_t = y_t - b_1 x_{1t}$, showing that, in this case, the two equations really have only one variable (x_0) in common. But in the first case, with b_1 unknown, it is different. The Jacobian of the second equation, applied to three points, vanishes if the first equation is taken into account; thus b_0 and b_1 cannot be found (though a_0 can). I wonder whether my error consists in wrongly applying the test or in wrongly maximizing the likelihoods. But it must be either the one or the other.

Question B). Your letter of July 2nd contains an additional (or rather alternative) condition under which prediction by least squares is legitimate; this in addition to the one developed in your letter of June 7th. I wonder how the two conditions apply to the two cases treated in the attached note. Would you say that in these cases all the parameters are unknown (condition one of your letter of July 2) because, though b_1 is known in the second case, this merely means that a new variable z_t (see above) is introduced. The "prediction" formula for both cases is given in (6.1) and is, naturally, a function of the exogenous variable x_1.

Sincerely yours,
J. Marschak

Annexed note: Maximum Likelihood Estimates of Parameters of a Pair of Random Equations
[p.1]
1. Consider the equations linear in the coefficients and valid for any time-point *t*.

$$(1.1a) \quad y_t = a_0 x_{0t} + \varepsilon_{1t}$$

$$t = 1, \ldots, N$$

$$(1.1b) \quad y_t = b_0 x_{0t} + b_1 x_{1t} + \varepsilon_{2t}$$

where ε_{1t} and ε_{2t} are the values of two independent normally distributed random variables with zero means, ε_1 and ε_2. y_t, x_{0t}, x_{1t} are observed values of y, x_0, x_1, measured from their means. Of these three variables, x_1 is "exogenous" with respect to the equations (1.1), in the sense that its value at any point of time is determined by some other, not specified, equations which do not involve x_0 and y. The latter two, on the contrary, are "endogenous" with respect to the equations (1.1) and do not occur in any other equations. Thus from the point of views of the system (1.1), x_{1t} is a constant; while y_t and x_{0t} are values of two random variables y and x_0; these values depend on x_{1t} and on the values assumed by the random variables ε_1 and ε_2, in the following way:

$$(1.2) \quad x_{0t} = \frac{b_1 x_{1t} + \varepsilon_{2t} - \varepsilon_{1t}}{a_0 - b_0}, \quad y_t = \frac{(b_1 x_{1t} + \varepsilon_{2t}) a_0 - \varepsilon_{1t} b_0}{a_0 - b_0}$$

[30] Section 19 of Haavelmo (1941) corresponds to section 19 of Haavelmo (1944).

as seen by solving (1.1). The problem is twofold: 1) to estimate the values of the coefficients a_0, b_0, b_1; 2) to predict y for given x_0 and x_1.
2. Economic example. Of the above two equations, the first one can be exemplified by the supply function, the second one by the demand function. Demand (equal to supply) is the variable y: the suppliers change their supply in response to changes in the commodity price x_0; the buyers change their demand in response to changes in the commodity price x_0 and in the "exogenous" variable, income, x_1.
[p.2]
3. The system can be generalized (dropping the time subscripts for brevity) into

$$y = a_0 x_0 + \Sigma_1^n a_i x_i + \varepsilon_1$$

$$y = b_0 x_0 + \Sigma_1^n b_i x_i + b_{n+1} x_{n+1} + \varepsilon_2$$

where the x's with non-zero subscripts are all "exogenous". (In terms of the previous economic example, $x_2, x_3, ...$ may be prices of other commodities). It is essential that the equations should not consist of similar terms only; otherwise the system becomes indeterminate in its coefficients (the case of multi-collinearity).
4. We shall be concerned here with the simple system (1.1) only. Two cases will be distinguished: 1) b_1 is unknown; 2) b_1 is known. (The second case arises, in our economic example when, in addition to time series of consumption, price and income, we can use family budgets to determine the income-elasticity of demand).
5. We have to derive the joint distribution of x_0, y from the joint distribution of $\varepsilon_1, \varepsilon_2$, using the 2$N$ equations (1.1) as transformation equations. The numerical value of the Jacobian

$$\frac{\partial(\varepsilon_{11}, ...\varepsilon_{1N}, \varepsilon_{21}, ...\varepsilon_{2N})}{\partial(y_1, ...y_N, x_{01}, ...x_{0N})} = \begin{vmatrix} 1 & 1 \\ -a_0 & -b_0 \end{vmatrix}^N = (a_0 - b_0)^N \quad ,$$

the latter expression being set as positive (in our example, $a_0 > 0, b_0 < 0$).

The distribution density (or "likelihood") of (x_0, y) is, accordingly,

$$(a_0 - b_0)^N \frac{1}{(2\sigma_1^2 \sigma_2^2 \pi)^{N/2}} \exp(-\frac{\Sigma\varepsilon_1^2}{2\sigma_1^2} - \frac{\Sigma\varepsilon_2^2}{2\sigma_2^2}),$$

where σ_1^2 and σ_2^2 are the variables of ε_1 and ε_2. Its logarithm is
[p.3]

$$L = -\frac{N}{2}\log 2\pi + N\log(a_0 - b_0) - \frac{N}{2}(\log\sigma_1^2 + \log\sigma_2^2) - \frac{\Sigma\varepsilon_1^2}{2\sigma_1^2} - \frac{\Sigma\varepsilon_2^2}{2\sigma_1^2}.$$

Maximizing L with respect to the parameters $\sigma_1^2, \sigma_2^2, a_0, b_0, b_1$, we obtain involving their maximum likelihood estimates $\hat{\sigma}_1^2, \hat{\sigma}_2^2, \hat{a}_0, \hat{b}_0, \hat{b}_1$, and expressions in terms of sumsquares and sumproducts of observations. Denote the latter as follows:
$\frac{1}{N}\Sigma y^2 = s; \frac{1}{N}\Sigma y x_0 = g_{0;} \frac{1}{N}\Sigma y x_1 = g_1; \frac{1}{N}\Sigma x_0^2 = h_{00;} \frac{1}{N}\Sigma x_1^2 = h_{11}; \frac{1}{N}\Sigma x_0 x_1 = h_{01}$.
We obtain [details omitted]

$$\hat{\sigma}_1^2 = s + \hat{a}_0^2 h_{00} - 2\hat{a}_0 g_0; \quad \hat{\sigma}_2^2 = s + \hat{b}_0^2 h_{00} + \hat{b}_1^2 h_{11} - 2\hat{b}_0 g_0 - 2\hat{b}_1 g_1 + 2\hat{b}_0 \hat{b}_1 h_{01};$$

(5.1) $$\frac{1}{\hat{a}_0 - \hat{b}_0} + \frac{g_0 - \hat{a}_0 h_{00}}{\hat{\sigma}_1^2} = 0; \frac{1}{\hat{b}_0 - \hat{a}_0} + \frac{g_0 - \hat{b}_0 h_{00} - \hat{b}_1 h_{01}}{h_{11}};$$

$$g_1 - \hat{b}_0 h_{01} - \hat{b}_1 h_{11} = 0.$$

Hence

$$\hat{a}_0 = \frac{g_1}{h_{01}}, \hat{b}_0 = \frac{s - \hat{a}_0 g_0}{g_0 - \hat{a}_0 h_{00}}, \quad \hat{b}_1 = \frac{g_1 - \hat{b}_0 h_{01}}{h_{11}},$$

$$\hat{\sigma}_1^2 = s + \hat{a}_0^2 h_{00} - 2a_0 g_0, \quad \hat{\sigma}_2^2 = s + \hat{b}_0^2 h_{00} + \hat{b}_1^2 h_{11} - 2\hat{b}_0 g_0 - 2\hat{b}_1 g_1 + 2\hat{b}_0 \hat{b}_1 h_{01.}$$

This gives the values of the 5 unknown parameters:

6. To predict *y* for given x_0 and x_1, we have to find the regression coefficient of *y* on x_0 (x_1 being given), $\frac{\Sigma x_0 y}{\Sigma x^2}$; remembering that ε_1 and ε_2 are independent we obtain from (1.2) the estimate

(6.1) $$E(y \mid x_0) = \frac{\Sigma x_0 y}{\Sigma x_0^2} = \frac{\hat{a}_0 \hat{b}_1^2 x_1^2 + \hat{a}_0 \hat{\sigma}_2^2 + \hat{b}_0 \hat{\sigma}_1^2}{\hat{b}_1^2 x_1^2 + \hat{\sigma}_1^2 + \hat{\sigma}_2^2} \cdot x_0$$

into which the values of the five parameters obtained in the preceding section are to be inserted.

[p.4]

7. The smaller σ_1 the more will the results be proximate to those obtained by multiple regression.

8. Consider now the case (mentioned under 4) where b_1 is known. The fifth of the equations (5.1) – the case obtained by maximizing with respect to b_1 drops out, while b_1 becomes a constant. For the other four parameters, we have [details omitted]

$$\hat{\sigma}_1^2 = s + \hat{a}_0^2 h_{00} - 2\hat{a}_0 g_0; \quad \hat{\sigma}_2^2 = s + \hat{b}_0^2 h_{00} + \hat{b}_1^2 h_{11} - 2\hat{b}_0 g_0 - 2\hat{b}_1 g_1 + 2\hat{b}_0 \hat{b}_1 h_{01};$$

$$\hat{a}_0 + \hat{b}_0 = \frac{2g_1 - \hat{b}_1 h_{11}}{h_{01}} = M \text{ (say)}$$

$$\hat{a}_0 \hat{b}_0 = \frac{M g_0 - s}{h_{00}} = N \text{ (say).}$$

From the last two equations,

$$\hat{a}_0 \text{ or } \hat{b}_0 = +\frac{M}{2} \pm \sqrt{\frac{M^2}{4} - N}.$$

It will be remembered that $a_0 > 0, b_0 < 0$, - this may help to choose the roots appropriately. From a_0 and b_0 , σ_1^2 and σ_2^2 are obtained by insertion.

9. Equation (6.1) can again be applied to find the prediction formula.

TH to JM: New York, July 9, 1943

Dear Marschak:

Thank you for your letter of July 8. I have not had much time to study the details of your calculations yet, so I cannot guarantee that they are all correct (but they probably are). I am, nevertheless, writing back at once because I thought that there are certain observations I could make which might to be of help for your work, and which I could make offhand from reading through your attached notes.

Concerning question A) in your letter

You mentioned here a test described in [Haavelmo, 1941] Section 19. From what you say on the top of the second page of your letter it occurs to me that perhaps you think that the content of Section 19 is something different from what it actually is. You mention a Jacobian in connection with the second equation in your example. My Section 19 does not deal with the equations describing an economic model. It deals with the joint probability distribution of the variables involved. The function f considered in my Section 19, although here no particular name is put on it, refers to a joint probability law of a set of variables. The whole Section 19 deals with the problem of finding general methods to investigate Problem I described on p.106-107.[31] The problem is, in general formulation, whether or not my equation (18.13) is satisfied, and that depends on the form of the function p (which in 19 I call f).

The test of determinacy of the parameters θ comes, roughly speaking, to this: Calculate the partial derivatives of the joint probability law with respect to each of the unknown parameters. This derivation will be functions of the variables as well as of the parameters. If these partial derivatives are linearly dependent (considered as functions of the variables), then in general it is not possible to estimate all the parameters. If the partial derivatives are not linearly dependent, estimation is in general possible.

From your remarks I get the impression that you are considering a much simpler test. If I am guessing correctly you proceed as follows:

You insert certain observations (2 points are enough, isn't it?) in the latter of your equations (1.1) to obtain equations that can be solved for b_0 and b_1. And then you do not get enough independent equations if you take the first equation into account. But then you are neglecting the effect of the error terms ε_1 and ε_2 (right?). I do not think that the determinant of your normal equations vanishes unless the variances of the error terms are zero. I think Frisch is responsible for the idea that the error terms do not matter for the idea of indeterminacy. But in general they matter very much. Now I do not know for certain whether the presence of error terms in your example is sufficient to insure determinacy. But if your maximum likelihood estimates are correctly calculated , and if they are unique (which I have not investigated) then there is no doubt that estimation of the parameters is possible. For one can give the following rule, which is an immediate implication of the rules I gave in Section 19 of [Haavelmo, 1941]:

If the maximum likelihood estimates of all the unknown parameters exist and are unique there cannot be any trouble with indeterminacy. The existence and uniqueness of the maximum likelihood estimates is a sufficient condition for determinacy. (For, obviously, the maximum likelihood estimates cannot be unique unless the partial derivatives of the probability law with respect to the parameters are linearly independent).

As I said, I have not had time to investigate whether your maximum likelihood estimates are actually unique. The formulae look all right from one point of view at least, namely that if x_1 is identically $= 0$ the estimates break down, which is as it should be.

Now concerning your question B)

I do not seem to be able to grasp what is the trouble with this point. However, I will try to repeat what I have already said, in a different way.

All I have tried to state is the simple mathematical proposition that if one has a function of several variables (the unknown parameters in the case of a likelihood function) one can find the maximum of this function in two ways, namely

[31] The passage in Haavelmo (1941) referred to is practically identical to "*I. The problem of confluent relations (or, the problem of arbitrary parameters)*" in Haavelmo (1944, p.91).

1. By maximizing with respect to the variables involved,

or,

2. by first transforming the variables into new variables, maximizing with respect to these new variables and inserting the values of the new variables thus found into the transformation equations to get the max values of the old variables. All provided the transformation is a one-to-one.

Now let us apply this to your particular case with a normal distribution. Your system (1.1) leads to the distribution of *y* and x_0 which you give at the bottom of page 2 in your note. It involves the "original" parameters $a_0, b_0, b_1, \sigma_1^2, \sigma_2^2$ (and the known parameters x_1).

Applying the maximum likelihood method to the corresponding likelihood function you obtain max. likelihood estimates of the 5 unknown parameters.

But you can also do something else. You can rewrite the distribution of the y's and x_0's in various forms, introducing new parameters that will be functions of the 5 original ones. In particular you can do the following: For every point of time *t* you can write the point distribution of *y* and x_0 on the form: The distribution of *y* given x_0 (and x_1) multiplied by the distribution of x_0 alone (given x_1). And the joint distribution of the *N* time points is the product of such distributions. In this distribution the original 5 parameters will occur in certain more or less complicated combinations. But you can simplify the form by introducing new parameters that are functions of the old ones. For example in the first part of the exponent of *e* you get an expression which can be written $(y_t - (Ax_{0t} + Bx_{1t}))^2$, where *A* and *B* are functions of the original 5 parameters $a_0, b_0, \sigma_1^2,$ etc. $(Ax_{0t} + Bx_{1t})$ is the expected value of y_t given x_{0t} (and x_{1t}). You can introduce 3 more new parameters, say *C*, *D*, *E*, so that the distribution will contain only the 5 new parameters A, B, C, D, E, and none of the old parameters. Assume now that there is a one-to-one correspondence between the old parameters $a_0, b_0, \sigma_1^2,$ etc. Then we have the following rule:

If you derive the max. likelihood estimates of *A*, *B*, etc. directly from the new form of the distribution you get exactly the same values as if you first derive the max. likelihood estimates of $a_0, b_0, \sigma_1^2,$ etc. from the old form of the distribution and thereafter insert these values in the equations between the old and the new parameters and calculate *A*, *B*, etc. from these equations.

If the relation between the old and the new parameters is not one-to-one, this rule is not correct.

Now, where does the least squares method come into the picture? It comes in by a coincidence so to speak, namely because, for the normal distribution it so happens that the least squares method leads to the same algebra as that which goes through in calculating max. likelihood estimates, as far as the estimation of the expected value of one variable given the others is concerned. This statement about the identity of the max. likelihood estimates of *A* and *B* and the least squares estimates of *A* and *B* (forgetting now for a moment that *A* and *B* are derived from $a_0, b_0, \sigma_1^2,$ etc.) is however not always true. Suppose for example that one would know that *C*, *D* and *E* were certain given functions of *A* and *B*. Then I think that the max. likelihood estimates of *A* and *B* taking account of the relation mentioned would not be the same as the straightforward least squares estimates of *A* and *B*.

Thus there are two problems to be considered, namely:

1. Are the "direct" maximum likelihood estimates of *A*, *B*, etc. the same as the "indirect" maximum likelihood estimates of *A*, *B*, etc., obtained by first estimating $a_0, b_0, \sigma_1^2,$ etc.

from the old form of the distribution and then calculating *A*, *B*, etc. from the transformed equations?

2. If one has a normal distribution involving certain parameters, to what extent does the max. likelihood method lead to the same algebra as the one involved in a certain least squares procedure?

Let me now recapitulate. Suppose that you have rewritten your distribution of the observable variables and introduced new parameters *A*, *B*, etc. corresponding to those which you want to estimate. And suppose that you accept the principle of maximum likelihood as a "good" principle of estimation. Then there are the following considerations to be made:

a) Are the "direct" max. likelihood estimates of *A*, *B*, etc. obtained from the new form of the distribution unique? If so, the problem of estimating *A*, *B*, etc. is solved.

b) Whether or not *A* and *B* can also be estimated (identically) by a least squares procedure is I think a rather trivial matter. One can find that out in each case.

c) If the "direct" max. likelihood estimates of *A*, *B*, etc. are unique and if there is a one-to-one correspondence between the old parameters a_0, b_0, σ_1^2, etc. and the new parameters *A*, *B*, etc. one can also obtain the max. likelihood estimates of *A*, *B*, etc. indirectly by first finding the max. likelihood estimates of the old parameters a_0, b_0, σ_1^2, etc.

d) If there are more new parameters *A*, *B*, etc. than old parameters a_0, b_0, σ_1^2, etc. one is wasting information by estimating all the new parameters independently by the direct max. likelihood estimation of them.

e) If there are fewer new parameters *A*, *B*, etc. than old parameters a_0, b_0, σ_1^2, etc. it shows that some of the old parameters can be chosen arbitrarily. There is indeterminacy of the old parameters with respect to the joint distribution of the observable variables.

Sincerely,
Trygve Haavelmo

Acknowledgements

This article draws on correspondence and documents from the Haavelmo Archive at the Department of Economics, University of Oslo, as organized by Tore Thonstad. I am most grateful to Wendy Glickman of the Rockefeller Archive Center for assistance in tracing Haavelmo's movements 1940-42. I am also most grateful for clarifying communication from Kenneth J. Arrow, J. J. Polak, and Paul A. Samuelson.

References

Arrow, K., 1978: Jacob Marschak, *Challenge*, March-April 1978, 69-71.

Bjerkholt, O., 2005: Frisch's econometric laboratory and the rise of Trygve Haavelmo's Probability Approach. *Econometric Theory* 21, 491-533.

Bjerkholt, O., 2007: Writing "The Probability Approach" with nowhere to go: Haavelmo in the United States, 1939-1944, *Econometric Theory* 23, 775-837.

Christ, Carl F., 1994: The Cowles Commission's Contributions to Econometrics at Chicago, *Journal of Economic Literature* 32, No.1, 30-59.

Cowles Commission (1952) *Economic Theory and Measurement: A Twenty Year Research report 1932-1952*. Cowles Commission, University of Chicago.

Frisch, R., 1933: *Pitfalls in the Statistical Construction of Demand and Supply Curves*. Veröffentlichungen der Frankfurter Gesellschaft für Konjunkturforschung, Neue Folge Heft 5, Leipzig: Hans Buske Verlag.
Frisch, R., 1938: Statistical versus theoretical relations in economic macrodynamics, Memorandum prepared the Business Cycle Conference at Cambridge, England, July 1938, to discuss Professor J. Tinbergen's publications for the League of Nations.
Haavelmo, T., 1938: The Method of Supplementary Confluent Relations, Illustrated by a Study of Stock Prices, *Econometrica* 6, No. 3, July 1938, 203-218.
Haavelmo, T., 1941: On the Theory and Measurement of Economic Relations, Mimeographed, Harvard University, Cambridge MA.
Haavelmo, T., 1943: The Statistical Implications of a System of Simultaneous Equations, *Econometrica* 11, No.1, 1-12.
Haavelmo, T., 1944: The Probability Approach in Econometrics, *Econometrica* 12, Supplement, 1-118.
Hendry, D. F. & M. S. Morgan, 1995: *The Foundations of Econometric Analysis*, Cambridge UK: Cambridge University Press.
Klein, L, 1991: The Statistics Seminar, MIT, 1942-1943, *Statistical Science* 6, No.4, 320-330.
Koopmans, T.J., 1945: Statistical Estimation of Simultaneous Economic Relations, *Journal of the American Statistical Association* 40, 448-466.
Mann, H. B. and A. Wald, 1943: On the Statistical Treatment of Linear Stochastic Difference Equations, *Econometrica* 11, 173-220.
Marschak, J., 1942: "Economic Interdependence and Statistical Analysis" in Lange, O., F. McIntyre and T. O. Yntema (eds.): *Studies in Mathematical Economics and Econometrics-In Memory of Henry Schultz*, Chicago.
Marschak, J., 1950: "Statistical Inference in Economics: An Introduction" in Koopmans,, T.C. (ed.): *Statistical Inference in Dynamic Economic Models*, Cowles Commission Monograph, No. 10, New York, 1-52.
Marschak, J. and W.H. Andrews, 1944: Random simultaneous equations and the theory of production, *Econometrica* 12, 143-172.
Morgan, M. S., 1990: *The History of Econometric Ideas*, Cambridge, UK: Cambridge University Press.
Qin, D., 1993: *The Formation of Econometrics*, Oxford: Oxford University Press.
Reder, Melvin W., 1982: Chicago Economics: Permanence and Change, *Journal of Economic Literature* 20, No.1, 1-38.
Spanos, A., 1989: On rereading Haavelmo: a retrospective view of econometric modeling, *Econometric Theory*, 5, 405-429.
Wilson, E.B, 1946: Review of T. Haavelmo: The Probability Approach in Econometrics, *Review of Economic Statistics* 28,173-174.