Koopmans 72

Amsterdam, Febr.21-th, 1936.

Professor R.Frisch,
Slemdalsveien 98,
Oslo.

Dear Professor Frisch,

I waited for some time before writing again to you because I was engaged in trying a new set-up of the same problem, which perhaps meets some of your objections against the use of sampling theory in regression analysis of economic data. Now I have a lot of things to write and of questions to put.

You will remember that I worked on the "universe" :

(1)
$$\begin{cases} x_k = y_k + \sigma \varepsilon_k z_k, \quad k=1,2\ldots K. \\ y_p = \sum_{m+1}^{K} \beta_r^{(p)} y_r + \beta^{(p)}, \quad p=1,2 \ldots m < K \\ \text{all } z_k \text{ mutually and of the } y_r \text{ independently normally distributed with mean } 0 \text{ and S.D. } 1, \\ \text{all } y_r,\ r=m+1\ldots K, \text{ normally correlated with moment matrix } \mu^*_{rs}. \end{cases}$$

As a first step assuming the ε_k to be known (they indicate the ratio's of the "disturbing intensities"), it may be proved that the values $\hat{\beta}_r^{(p)}, \hat{\beta}^{(p)}, \hat{\mu}^*_{rs}, \hat{\sigma}$ of maximum likelihood of the parameters in (1), deduced from a sample $x_k^{(t)}$, $t=1\ldots T$, $k=1\ldots K$, are such that the vectors in the rows of

$$\begin{matrix} -1 & 0 & \cdots & 0 & \hat{\beta}^{(1)}_{m+1} & \cdots & \hat{\beta}^{(1)}_K \\ 0 & -1 & \cdots & 0 & \hat{\beta}^{(2)}_{m+1} & \cdots & \hat{\beta}^{(2)}_K \\ \cdot & \cdot & \cdots & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & \cdots & -1 & \hat{\beta}^{(m)}_{m+1} & \cdots & \hat{\beta}^{(m)}_K \end{matrix}$$

are linear combinations of the m characteristic vectors corresponding to the m smallest characteristic roots of $E^{-2}\mathcal{M}$, where

$$E = \left\| \begin{matrix} \varepsilon_1 & 0 & \cdots & 0 \\ 0 & \varepsilon_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \varepsilon_K \end{matrix} \right\|$$

and $\mathcal{M}$ = sample moment matrix.

This means, that, after applying a stretch which makes all $\varepsilon_k = 1$, we have the orthogonal $(K-m)$-dimensional regression hyperplane.

A serious objection against the applicability of this result is that the assumption about the distribution of the $y_{m+1}\ldots y_K$ - the independent ones of the systematic components - is of a very special character, and can hardly be expected to be satisfied in any economic application. And it did not seem possible to extend the above theorem to more general distributions of the $y_{m+1}\ldots y_K$.

This led me to try a new set-up in which every $y_{m+1}\ldots y_K$, for every moment of observation $t=1\ldots T$, itself is considered as an unknown parameter $\eta_k^{(t)}$. The system now becomes for $m=1$:

~~x - - z~~

Ragnar Frisch

$$(2)\quad \begin{cases} x_k^{(t)} = \eta_k^{(t)} + \sigma \varepsilon_k z_k^{(t)} & k=1\ldots K,\ t=1\ldots T \\ \beta_\kappa \eta_\kappa^{(t)} + \beta = 0 & \text{(summation over any greek index occuring at least twice)} \\ z_k^{(t)} \text{ as in (1)} \end{cases}$$

This is not a "universe" in the usual sense; in fact every new observation $x_k^{(t)}$, $k=1\ldots K$, introduces $K-1$ new unknown parameters, for instance $\eta_2^{(t)}, \eta_3^{(t)} \ldots \eta_K^{(t)}$ (if $\beta_1 \neq 0$). Consequently the accuracy of estimation of the $\eta_k^{(t)}$ must be very poor, and cannot be increased beyond any limit by increasing T, the number of observations. However, we do not need the $\eta_k^{(t)}$, but only the β_k and β, and these may indeed be estimated with an accuracy which surmounts any limit if $T \to \infty$, provided that the (singular) moment matrix $\mu_{k\ell}$ of the $\eta_k^{(t)}$ tends to some limit.

As far as now I only worked out this set up for the case $m=1$, or the fitting of one single regression equation. It is not certain whether maximum likelihood estimation for this unusual universe has the same optimum properties which Fisher proved for the usual universe. The formal procedure, however, seems in this case to lead to something which is useful. The solution $\overset{*}{\beta}_k$ (further on I will write b_k) is, for given ε_k (which in maximizing the likelihood function are taken as constants), exactly the same as in the case (1) : the orthogonal regression after application of a stretch which makes all ε_k equal. In the limiting case $\varepsilon_1=1$, $\varepsilon_2=\varepsilon_3=\ldots=\varepsilon_K=0$, this is the first elementary regression.

It is clear that all results, deduced from this set-up, are independent of the time-shape of the $\eta_k^{(t)}$-curves. In fact, the only characteristics of these curves, entering into the sampling distribution of the b are the β_k and the moments of the $\eta_k^{(t)}$, in first approximation only the second degree moments $\mu_{k\ell}$. Therefore, the hypothesis of independent normal distribution in consecutive observations is now retained only with regard to to the "disturbances" $\sigma \varepsilon_k z_k$ in the variables, which widens the field of application considerably, and may be controlled in any practical case.

An interesting point is that this scheme permits a combined treatment of what we called "Sampling approach" and "Limit approach". There are two possible sources of divergence between a regression b_k computed from the sample and the "universe" regression β_k :

a/ The "sampling effect": b_k may differ from its mathematical expectation $E\,b_k$, and

b/ the "Uncertainty effect": the assumed values ε_k^*, indicating the ratio's of the "disturbing intensities", which have been used in computing b_k from the sample, may have been chosen different from the unknown universe values ε_k, which leads to a divergence between $E\,b_k$ and β_k.

The second effect corresponds to the property shown on p.60 of your book: the elementary regressions as limits of the "true" one, if the sampling effect is neglected. The first ~~xxxxx~~ effect is the only one which is considered in the classical sampling theory of regression. It is of interest to compare the order of magnitude of both effects.

The sample moment matrix $m_{k\ell}$, as computed from the observations (2), may be developed with regard to σ:

(3) $m_{k\ell} = \mu_{k\ell} + \sigma m'_{k\ell} + \sigma^2 m''_{k\ell}$,

where $\mu_{k\ell}$ is the (singular) moment matrix of the $\eta_k^{(t)}$, and $m'_{k\ell}$ and $m''_{k\ell}$ are homogeneous expressions of first resp. second degree in the $z_k^{(t)}$. Now in the definition equation of the b_k

(4) $(m_{k\lambda} - \ell \varepsilon_k^{*2} \delta_{k\lambda}) b_\lambda = 0$ (ℓ smallest char. value, $\delta_{k\ell} = 1,\ k=\ell;\ 0,\ k\neq\ell$)

we may develop b_k in a similar series

(5) $b_k = \beta_k + \sigma b'_k + \sigma^2 b''_k + \ldots\ldots$,

where b'_k , $b''_k \ldots$ are defined by the requirement that the coefficients of any power of σ in (4) should equal zero (analogously to the perturbation theory in quantum mechanics).

The first term in this development, b'_k , represents the "sampling effect" in first approximation. It may be shown that the variables b'_k , $k=1\ldots K$, are normally distributed about 0 with moment matrix

(6) $\pi_{k\ell} = \frac{\sigma^2}{T} (\varepsilon_\lambda^2 \beta_\lambda^2)(\mu^{-1}_{k\ell} + \phi_k \beta_\ell + \beta_k \phi_\ell)$, ϕ_k arbitrary .

Here $\mu^{-1}_{k\ell}$ represents the matrix which, by the orthogonal transformation which brings $\mu_{k\ell}$ to the form $\left\| \begin{matrix} 0 & 0 & \cdots & 0 \\ 0 & \mu_2 & \cdots & 0 \\ \vdots & \vdots & & \\ 0 & 0 & \cdots & \mu_K \end{matrix} \right\|$ itself takes the form $\left\| \begin{matrix} 0 & 0 & \cdots & 0 \\ 0 & 1/\mu_2 & \cdots & 0 \\ \vdots & \vdots & & \\ 0 & 0 & \cdots & 1/\mu_K \end{matrix} \right\|$

and therefore could be called a partial inverse of $\mu_{k\ell}$. Further, the arbitrary summand $\phi_k \beta_\ell + \beta_k \phi_\ell$ is due to the fact that the b'_k by the nature of their definition contain an arbitrary summand $\phi \beta_k$. The factor $\varepsilon_\lambda^2 \beta_\lambda^2$ in (6) may be discarded by taking $\varepsilon_\lambda^2 \beta_\lambda^2 = 1$ as the rule which normalizes β_k. Now, taking $\phi_k = 0$, (6) shows that $b_k - Eb_k$ is at most of order of magnitude

(7) $\frac{\sigma}{\sqrt{\mu_2 T}}$ $(\mu_2 < \mu_3 < \cdots < \mu_K)$

and is in first approx. quite independent of the discrepancy between assumed (ε_k^*) and true (ε_k) "disturbing intensities". Another special choice of ϕ_k, which makes $\pi_{11} = 0$, leads to the usual formula for the sampling distribution of the coefficients of the first elementary regression, which is therefore contained in (6) as a special case. In its general form, however, (6) does not show the defaults which you critisized in the special form, namely, that it does not detect the dangerous cases in which more than one linear relations are approximately satisfied in the material.

For, in such cases, μ_2 must be small compared with μ_K, and this must in most of the samples manifest itself too in the sample moment matrix $m_{k\ell}$ (I am aware that this formulation is not invariant for a stretch; it seems to me, however, that the point that a test built on (6) may detect multiple collinearity is independent of the units of measurement of the x_k).

Therefore I hold that the default of the usual regression analysis of not detecting the uncertainty in the results due to multiple collinearity in the material is a consequence of the asymmetric treatment of the variables.

The second term in (5), $\sigma^2 b''_k$, reveals the "uncertainty effect", or the bias due to a wrong choice ε_k^* of the ratios of the disturbing intensities $\sigma\varepsilon_k$. The mathematical expectation of $\sigma^2 b_k''$ is :

$$(8) \qquad \sigma^2 Eb_k'' = \sigma^2\left(1-\frac{K-1}{T}\right)(\varepsilon_\lambda^2\beta_\lambda^2)\left(\frac{\mu^{-1}_{kk}\varepsilon_k^{*2}\beta_k}{\varepsilon_\nu^{*2}\beta_\nu^2} - \frac{\mu^{-1}_{kk}\varepsilon_k^2\beta_k}{\varepsilon_\nu^2\beta_\nu^2}\right) + \text{terms of order } \frac{\sigma^2}{\mu_2 T}$$

The main term is zero for $\varepsilon_k^* = \varepsilon_k$. If this is not the case, $Eb_k - \beta_k$ has the order of magnitude

$$(9) \qquad \frac{\sigma^2}{\mu_2} .$$

The uncertainty effect is therefore, as could be expected, independent of the number of observations T . However, with respect to σ , the common factor in all disturbance intensities, it is of one order higher compared with the sampling effect (7). Hence it depends upon the circumstances of the problem which of both effects is responsible for the main part in the divergence $b_k - \beta_k$, and, I think, both effects deserve attention.

For the case $\varepsilon_1^*=1$, $\varepsilon_2^* = \varepsilon_3^* = \ldots\ldots = \varepsilon_4^* = 0$, is the expression (8) closely related to the first elementary regression in the universe (2). This point is not yet quite clear to me because of a slight discrepancy in the numerical factor.

The use of the orthogonal regression b_k after a stretch given by ε_k^* is a point in common with Hotelling's method, and the question arises in how far this set-up may avoid the traits of that method which you critisized in your letter of Dec.14 as being inadequate to the requirements of an analysis of economic data. I hold it may.

In the first place, here the determination of the smallest principal axis after a stretch is not intended as a means of finding a set of "basic" variables, but only as a means of finding an estimate of the universe regression β_k. Now the ~~assumption of given~~ choice of assumed values ε_k^* for the ratios of the disturbing intensities, which supplies the metric, necessary to give any sense to an orthogonal regression, seems at first sight to be as artificial as Hotelling's requirement that the independent basic variables should be found by an orthogonal transformation. However, we can use our formulae to study how the ~~solution depends~~ bias in our solution, due to the difference $\varepsilon_k^* - \varepsilon_k$, depends on that difference, and (8) supplies in first approximation the means to that.

Secondly, the testing of superfluity may be treated here simply as a test on the significance of the deviation from zero of a regression coefficient. As such a regression coefficient is always determined for a set of variables this satisfies your postulate that the concept of superfluity should refer to a variable as a member of a set of variables.

I have planned further work on the following points :

1. Can remaining terms in the development of b_k be neglected ? Criteria to answer this question. If this is the case, the first logical step has been made : Given the universe, how ~~is the~~ are the statistics which are intended to estimate its parameters distributed ? Remains the inversion : Given the sample, what can be said about the universe? This leads to :

2. Construct from (6) and (8) a test, deciding with a small risk of error, whether a given sample, ~~or better a sample with a given set of statistics b~~ may have been drawn from a universe with given β_k , whatever its $\mu_{kl}, \varepsilon_k, \sigma$.

Further points are :

3. How to test whether the hypothesis of independent succesive $z_k^{(t)}$-values is satisfied ?

4. What is the influence of this hypothesis not being exactly satisfied?

I will be very grateful if you would write to me your reaction on all this. There is one more point on which especially I would be very glad to have your comment. I am questioning whether it has any use to try to generalize these results to the problem of fitting at once more than one relation to a multiply collinear scatter diagram. I am aware of two different practical cases in which a material may be multiply collinear. The first case which I have in mind is a case in which there is between the variables only one economically significant relation, but where an "accidental" second relation deprives us of the opportunity to determine the first one. An example could be this :

x_1 income of a group of consumers
x_2 price of beaf
x_3 price of porc
x_4 consumption of beaf in the group .

There can be expected one economically significant relation

(11) $x_4 = ax_1 - bx_2 + cx_3$.

It could occur, however, by chance of crop variations or tarif policy, that a second relation, say

(12) $x_1 = dx_2$

happened to be approximately satisfied in the period concerned. In this case the problem is not to determine both relations, but only to detect in how far the accuracy of determination of a, b, c is vitiated by the multiple collinearity in the material. I expect that a test built on (6) and (8) will enable us to deal with this problem.

A second possibility is, that indeed more than one economically significant relation exists between the variables. An example is obtained if we add as fifth variable

x_5 consumption of porc in the group.

Now we may expect, besides (11), a relation

(13) $x_5 = ex_1 + fx_2 - gx_3$,

together 2 relations between 5 variables. Mathematically they are equivalent to any two independent combinations of them. The economically significant ones, however, are obtained by taking those two equations out of the whole two-dimensional set for which the coefficients of x_4 resp. x_5 are zero. Now, if no accidental relations of the type (12) disturb the matter, how to fit these two relations ? In your book a symmetric procedure is given : fitting one relation ~~txx~~ to every subset of 4 variables, and determining from these 5 relations two resultant ones by compatibility smoothing. But would it, in this case, not be the more natural thing thing, to fit only two relations, one to the 1234- and one to the 1235-set ? Otherwise our estimate of the elasticity of beaf consumption would be influenced by the errors in the statistics of porc consumption, which is quite outside the problem.

In this example the economically significant relations in the big 12345-set are found by splitting up the set into two subsets 1234 and 1235, in each of which only one economically significant relation exists. Is such a splitting always possible, or are there problems in which two or more economically significant relations contain all variables out of the same set ? If such cases occur, a more symmetric treatment seems to me to be the natural thing.

There are still more points, but I think, now this letter is long enough. At any rate, you are informed now about the present stage of my work.

Every now and then, I feel a longing to the beauty of the Norwegian mountains and skies. At the other hand, I much enjoy here having bright daylight already from 8 till 18 o'clock.

In London, I promised Professor Pearson, Department of Applied Statistics, University College, Gower street, London W.C.1, to ask you to send to him a copy of your paper on substitution in the chocolate factory. He is very interested in industrial statistics. Would you be so kind to do so ?

Yours sincerely,

T. Koopmans

T.Koopmans,

address now :

Roerstraat 115III

Amsterdam Z.

COWLES COMMISSION
FOR RESEARCH IN ECONOMICS
THE UNIVERSITY OF CHICAGO
CHICAGO 37, ILLINOIS

May 20, 1948

Professor Ragnar Frisch
Vinderen
Oslo, Norway

Dear Professor Frisch:

I send kindly this book

Mr. Preben Munthe kindly showed me mimeographed notes of your lectures on economic theory given at the University in Oslo. The two volumes he showed me are marked fourth edition. We are much interested in this material, particularly the 17th lecture on the economic circulation system. I wonder if you would be willing to let us have a copy of each of these two volumes for incorporation in our working library, which is used by the Cowles Commission Research staff and graduate students, but not generally by the public.

If you have a spare copy available, we would also appreciate receiving your article in Ekonomisk Tidskrift, Uppsala, 1943.

With best personal regards,

Sincerely yours,

Tjalling C. Koopmans

Tjalling C. Koopmans

TCK:MMM

June 16, 1948.

Dr. Tjalling Koopmans,
Cowles Commission,
The University of Chicago,
Chicago 37,
Illinois,
U.S.A.

Dear Dr. Koopmans,

At the request of professor Frisch I am to-day sending you under separate cover the two volumes of "Notater til økonomisk teori" and his article in Ekonomisk Tidskrift "Økosirk-systemet", Uppsala 1943, which you mentioned in your letter of May 20.

Sincerely yours

Inger Østraat
Secretary to Professor Frisch

AIRMAIL

If you have a spare copy available, we would also appreciate receiving your article in Ekonomisk Tidskrift, Uppsala, 1943.

With best personal regards,

Sincerely yours,

Tjalling C. Koopmans

TCK:MMM
Tjalling C. Koopmans

COWLES COMMISSION
FOR RESEARCH IN ECONOMICS
THE UNIVERSITY OF CHICAGO
CHICAGO 37, ILLINOIS

October 13, 1948

Professor Ragnar Frisch
Vinderen
Oslo, Norway

Dear Professor Frisch:

I have not before this answered your very interesting circular letter of May 20, mainly because I wanted to see completed a manuscript on identification problems, and I enclose that with this letter as my modest contribution to the collection of materials and ideas that you are undertaking. The manuscript deals with one formal methodological problem that arises in the construction of economic models, and a problem to which you have in the past given much thought. This is the problem of distinguishing the identity of economic relationships before undertaking their measurement.

In the Cowles Commission we have during the past years attempted to formalize the problem of identification and to develop definite criteria of identifiability in linear models. In this expository manuscript I attempted to demonstrate the meaning of the notion of identifiability and the application of its criteria. You will recognize, perhaps in somewhat different terminology, many of the ideas that are expressed in your own work on the measurement of economic relationships, particularly your memorandum prepared in 1938 for the conference in Cambridge, England, on Tinbergen's work.

[illegible]

Incidentally, I intend to submit this article to Econometrica after I have had the benefit of comments from a few colleagues. It is not yet being submitted herewith, but I thought that it might be useful to you in the work for the United Nations Subcommission on Employment mentioned in your circular letter. I would be grateful for any comments you may care to make.

Sincerely yours,

Tjalling C. Koopmans

Tjalling C. Koopmans
Director of Research

TCK/fs
Enclosure

November 17, 1948.

Dr. Tjalling C.Koopmans,
Director of Research,
Cowles Commission for Research in Economics,
The University of Chicago,
Chicago 37,
Illinois,
U.S.A.

Dear Dr. Koopmans,

At the request of Professor Frisch I am writing to you to thank you for your letter of October 13, and your memorandum "Identification Problems in Economic Model Construction". It is taken as a memorandum in the University Institute of Economics' series of stencil-memo and will be distributed amongst others to the students. It will be mentioned that it is a "Cowles Commission Discussion Paper". Professor Frisch hopes that you have no objection as to this.

Sincerely yours

Inger Østraat
Secretary to Professor Frisch

AIRMAIL