20 Lectures on Introduction to Econometrics
Given at the University of Chicago
Spring, 1949

Cowles Commission for Research in Economics
The University of Chicago
Chicago 37, Illinois
INTRODUCTION TO ECONOMETRICS
20 Lectures given at the University of Chicago in Spring, 1949*

Part I. Non-stochastic economics

Lecture 1. Best policy. Goal variable; non-controlled, controlled, strategic variables.

2. Exogenous variables and structural parameters. Types of prediction.

3. Determining the structure from theory and data.

4. An example.

5, 6. Econometric "pitfalls" due to disregarded variables or relations. Non-identifiable structures.

7. The identification, continued.

8. Why does the identification problem arise in non-experimental sciences?

9, 10. Discussion of earlier problems.

11. When need we know the structure?

Part II. Stochastic economics: Population properties

12. Joint distributions, non-parametric.


14. Parameters of joint distributions.


16. Exogenous and endogenous variables in stochastic economics.

17. Identification and determination of structure by the method of reduced form: examples.

18. More examples.


Part III. Stochastic economics: Sample properties

20. Useful properties of certain least squares and maximum likelihood

*To be used jointly with 24 Lectures (same title) given at the University of Buffalo in Spring, 1948.
estimates. Obtaining maximum likelihood estimates of structure from those of reduced form.

---

Recommended reading:

Attached materials:

1. J. Marschak, "Economic Structure, Path, Policy and Prediction"

2. "Statistical Inference from Non-Experimental Observations—an Economic Example"

3. C. Hildreth, "Problems in the Estimation of Agricultural Production Functions"

**As far as available.**
ERRATA

Chicago Lectures

Page

1.3 In Exercise 1.3, p, x, z must be replaced everywhere by $p$, $x$, $z$.

1.4 Par. 3, line 1. Delete 1).

1.4 Par. 3, line 3. Replace "and" by "and/or".

2.1 Equation (2.5) Replace $1/2\theta$ by $(1/2)\theta$.

2.1 Equation (2.6) " $\theta/2$ by $\theta/2$.

Line 2 from bottom " 2.1 by 1.9.

4.2 Problem 4.1. Add the following new paragraph:

Using the figures for 1939 and 1946, one obtains the following approximate answer: $x = 36 - 0.04 p + 0.0114 y(1 - \tau)$, where $x$ is measured in billion gallons, $p$ in cents per 100 gallons (wholesale) and $y$ in billion dollars.

5.1 Par. 2, line 4. For "classified," read "clarified".

7.1 Par. labelled 110: after "necessary" insert "See 1110, 1111."

7.1 Exercise 8.1, line 2-3. After "suppose" the sentence should read as follows:

One has to estimate the combined effects 1) of changing income and 2) of introducing for the first time a sales tax.

11.3 For "Exercises 11.2 and 11.3" read "Exercises 11.2-11.3".
Lecture 1.

Knowledge is called useful if it helps to choose the best policy (action).

Exercise 1.1 The demand curve for the product of a firm in an imperfect market is known to be linear, with unknown parameters; the firm's total cost is known to be independent of output. Which are the parameters that (A) the firm has to know for determining its best policy? (B) the government has to know for determining the rate of an excise tax on the firm's product which will give the highest returns?

(A) Let \( x = \) amount sold, \( p = \) price, \( z = \) profit

\[(1.1) \quad x = c - \beta p \] (demand function)
\[(1.2) \quad z = px - \gamma \text{, where } \gamma = \text{total cost}.\]

Best policy: choose \( p \) (or choose \( x \)) so as to maximize \( z \).

Variables:
- Non-controlled: \( c, \beta, \gamma \)
- Controlled: \( x, p, z \)
- Goal variable: \( z \)
- Strategic variables: \( p, x \).

Number of controls \( (\text{= degrees of freedom}) \): 1

Notation for best values of controlled variables: \( \hat{x}, \hat{p}, \hat{z} \).

Maximization condition:
\[(1.3) \quad \frac{dz}{dx} \bigg|_{p = \hat{p}} = 0. \text{ (Read: derivative of } z \text{ with respect to } x \text{ vanishes when } p = \hat{p}).\]

\[(1.1), (1.2) \text{ yield }\]
\[(1.4) \quad z = p(x - \beta p) - \gamma ; \text{ and therefore (1.3) yields}\]
\[(1.5) \quad \hat{p} = c/2 \beta \]
\[(1.6) \quad \hat{x} = c/2 \]
\[(1.7) \quad \hat{z} = (c^2/4\beta) - \gamma .\]

Either (1.5) or (1.6) suffices to describe best policy. Hence, the answer to the question in (A): To determine best policy, the firm has to know \( c, \beta \) (but not \( \gamma \)).

Before proceeding to (B) (government policy) let us generalize problem (A) into (A') in which an excise tax of \( \Theta \) dollars per unit exists. [That is, (A) is a special case of (A'), with \( \Theta = 0 \).] We have to redefine profit by generalizing (1.2) into

\[(1.2') \quad z = (p - \Theta) x - \gamma ,\]

while (1.1), (1.3) remain valid. We now derive the best values of \( \hat{p}, \hat{x}, \hat{z} \) when the excise tax is \( \Theta \). Call them \( \hat{p}(\Theta), \hat{x}(\Theta), \hat{z}(\Theta) \):
\[ (1.5') \quad \hat{p}(0) = (a + \beta 0)/2 \beta \]
\[ (1.6') \quad \hat{x}(0) = (a - \beta 0)/2 \]

In this notation, the expressions in (1.5), (1.6) become \( \hat{p}(0), \hat{x}(0) \). Check!

One non-controlled variable has been added to the previous list, viz., \( \theta \). To determine best policy, as in (1.5') the firm has to know \( a, \beta, \theta \).

Problem (B) deals with the best policy of the government, not the firm, viz.: choose \( \theta \) so as to maximize the tax revenue \( T \), defined by

\[ (1.8) \quad T = \theta \cdot \hat{x}(0); \quad \text{or, because of (1.6')}, \]
\[ (1.9) \quad T = \theta \cdot (a - \beta 0)/2. \quad \text{The maximization condition is} \]
\[ (1.10) \quad \frac{dT}{d\theta} = 0, \quad \text{where} \quad \theta^* \quad \text{is the optimal tax rate (from the}
\text{government's point of view). Note that (1.9) is quite similar to (1.4) (with} \theta \text{ replacing } p; \quad \text{the constants } \gamma \text{ and } \alpha \text{ being irrelevant to the}
\text{maximization). We have} \]
\[ (1.11) \quad \theta^* = a/2 \beta. \]

That is, the government needs to know \( a, \beta \) in order to choose the best policy in the sense of maximum tax revenue. Summarizing the list of variables from the government's point of view:

Non-controlled: \( a, \beta, \gamma \) (but \( \gamma \) has no effect upon goal \( T \) or strategy \( \theta \)).

Controlled: \( \theta, T \)

Goal: \( T \)

Strategy: \( \theta \)

Degrees of freedom: 1

Such might be the case with the government by a conqueror. In a democracy, goals other than \( T \) are present. Example: the government wants to obtain high \( T \) as well as to induce high \( \hat{x} \). It maximizes a "utility function"—say, a linear one:

\[ (1.12) \quad U = T + \omega \hat{x}, \]

where the parameter, or weight, \( \omega > 0 \) indicates how many dollars in tax revenue the government is willing to sacrifice in order to increase the production by one unit.

Further exercises:

Exercise 1.2 Let \( x \) of the preceding exercise be measured in gallons, and \( z \) in dollars. Give the dimensions of \( a, \beta, \gamma, \theta \) and show that the right-hand sides of (1.5), (1.6), (1.7), (1.5'), (1.6'), (1.11) have the same dimensions.
as the respective left-hand sides.

**Exercise 1.3** Tabulate the results of 1.1 as follows

<table>
<thead>
<tr>
<th>θ = 0</th>
<th>θ &gt; 0</th>
<th>θ = θ* = c/2β</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Compute (and comment in words upon) p(θ) - p(0); x(θ) - x(0); P(θ*)/p(0); θ*(θ*)/x(0); θ*(θ*) - z(0); T (θ*)/z(θ*).

**Exercise 1.4** Does the check suggested after equation (1.6') guarantee against error? Using the symbols →, ←, ↔ for "sufficient," "necessary," "equivalent," write the appropriate relation between propositions (1.5) and (1.5').

**Exercise 1.5** Find the best values θ*, T*, U* when the goal is to maximize U in (1.12).

**Exercise 1.6** Suppose the firm systematically fails to maximize its profits, always charging a price that exceeds the optimal price by $Δ$ dollars per gallon. If tax revenue is maximized at θ*, express θ* as a function of c, β, γ, δ.

**Exercise 1.7** How would problem (1.1) and its solution be modified if total cost were dependent on output x?

**Exercise 1.8** Modify problem (1.1), and give solution, for the case when the market is competitive.
Exercise 1.9. Suppose the American demand for a foreign monopolist's product is a linear function of its price and of the disposable national money income; the latter is defined as the excess of the national money income over the income tax revenue. The income tax rate is assumed flat, i.e., independent of income. The monopolist's total cost and his revenue from non-American markets are assumed independent of American demand.

It is further assumed that human behavior is sufficiently stable, and that data permit the evaluation of constants characterizing any past relations between variables. For example, if any two (or three) observable variables are connected by a linear relation, all its coefficients can be evaluated from observations, i.e., a line (or plane) can be "fitted."

Question: 1) What theories (if any) and past data will you use and what constants will you evaluate in order to predict the effect of a given change in the tariff rate (per gallon of the monopolist's product), and of a given change in income tax rate, upon the total tax revenue, in each of the following four cases:

a) the excise tax rate had varied every previous year; the income tax rate was zero or constant;

b) the income tax rate had varied every previous year; the excise tax rate was zero or constant;

c) both tax rates had varied every previous year;

d) both tax rates were constant or zero every previous year.
Lectures 2 and 3.

Exercise 1.5 was discussed in detail. Retain notations of Lecture 1 and write \( y = \text{income}, \quad \tau = \text{income tax rate}, \quad \delta = \text{a constant}. \) Then

\[(2.1) \quad x = c = \beta y + \delta (1 - \tau)y.\]

That is, the "constant term" (or "intercept") \( c \) in (1.1) is now replaced by a more general expression, \( c \equiv \delta (1 - \tau)y. \) Equation (1.2) remains valid, with \( y \) redefined to mean "total cost of production minus total sales abroad." Equation (1.3) remains as it stands. We obtain therefore,

\[(2.2) \text{ instead of (1.5')}: \quad \hat{y}(\theta) = [c + \delta (1 - \tau)y + \beta \theta]/2 \beta ; \]
\[(2.3) \text{ instead of (1.6')}: \quad \hat{x}(\theta) = [c + \delta (1 - \tau)y - \beta \theta]/2 ; \]
\[(2.4) \text{ instead of (1.8'):} \quad \tau = \theta \hat{x}(\theta) + \tau y. \quad \text{And, substituting from (2.3),} \]
\[(2.5) \quad \tau = 1/2\theta [c + \delta (1 - \tau)y - \beta \theta] + \tau y \]
\[= \tau(\theta, \tau, y, c, \beta, \delta), \text{ say.} \]

The effects of small changes of \( \theta \) and \( \tau \) upon \( \tau \) are

\[(2.6) \quad \partial \tau/\partial \tau = y(1 - \theta/2), \text{ a function of } \gamma, \theta, y; \]
\[(2.7) \quad \partial \tau/\partial \theta = -\beta \theta + [c + \delta (1 - \tau)y]/2, \text{ a function of } y \text{ and all} \]
\[\text{Greek letters involved.} \]

As long as \( \partial \tau/\partial \tau \) and/or \( \partial \tau/\partial \theta \) is larger or smaller than zero, it is worth while to change the tax rates, i.e., they are not optimal.

Equating each of these derivatives to zero, we obtain two equations to determine the optimal tax rates \( \theta^*, \tau^* : \)

\[(2.8) \quad \theta^* = 2/\delta \]
\[(2.9) \quad \beta \theta^* = [c + \delta (1 - \tau^*)y]/2. \]

Thus \( \theta^* \) turns out to depend on \( \delta \) only; while \( \tau^* \) depends on \( c, \beta, \delta, \) and the income \( y. \)

In practice the equations (2.6), (2.7) are used not only to determine the optimal policies \( \theta^*, \tau^* \), but also to state the effect of any given change in policy; if \( \theta \) is changed by a small amount \( \Delta \theta \), and \( \tau \) is changed by a small amount \( \Delta \tau \), then the increment of the tax revenue,

\[(2.10) \quad \Delta \tau = (\partial \tau/\partial \theta) \cdot \Delta \theta + (\partial \tau/\partial \tau) \cdot \Delta \tau \text{ approximately, where the values of} \]
\[\partial \tau/\partial \theta, \partial \tau/\partial \tau \text{ are given by (2.6), (2.7).} \]

The question asked in Exercise 2.1 is: What measurements are necessary in order to predict the value of \( \Delta \tau \) in (2.10), for given values of \( \Delta \theta, \Delta \tau; \)
[or to predict the value of T in (2.5) for given values of , see also Exercise 2.9 below]. In particular the case (a) and (c) of Exercise 1.9 can be described by classifying the quantities involved as follows:

<table>
<thead>
<tr>
<th>Case (a)</th>
<th>Case (c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Have changed</td>
<td>Yes</td>
</tr>
<tr>
<td>Are going to change</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>θ, Y</td>
</tr>
<tr>
<td>No</td>
<td>β, δ</td>
</tr>
</tbody>
</table>

Cases (b) and (d) can be recorded similarly.

Case (c) has an important property. In this case, the data themselves can tell the nature of the relation between the "dependent variable" or "predictand" ΔT and the "independent variables" or "predictors" Δθ, Δτ, y, θ, τ [or the predictand T and the predictors θ, τ, y]. The method of "regression analysis" and its graphical counterpart (as described in Ezekiel's "Methods of Correlation Analysis") has been designed for just such a case. No theory is here needed besides the statement that T depends on certain specified variables. To be sure, theory helps to specify the relationship somewhat. For example, because of (2.6), (2.7) the function on the right-hand side of (2.10) has the form

\[(2.11) \quad ΔT = y Δτ + ε_1 Δθ + \varepsilon_2 + ε_3 Δθ + ε_4 (1 - τ) y Δθ,\]

where the ε's are certain constants which can be "fitted" to the time series of the following variables:

\[y Δτ, \theta y Δτ, \theta Δθ, Δθ, (1 - τ) y Δθ;\]

and once the are so "fitted," these ε's can be used to predict ΔT when new values of y, θ, τ, Δθ, Δτ are given. In principle, even this knowledge of theory is not needed: the data themselves will tell whether, for example, ΔT is a linear function of the relevant increments,

\[(2.12) \quad ΔT = π_1 Δτ + π_2 Δθ + π_0,\]

or whether ΔT is a non-linear function of these increments and of y, θ, τ, as happens to be the case in (2.11). If changes are small, (2.12) may be good enough as a first approximation, even though it contradicts the "theoretical" relationships (2.6), (2.7) derived from the assumption that the monopolist maximizes his profit.

The fitting of (2.12), or of some other simple relation between T and the "predictors," by computation or by graphical method, is, in fact, in line with much of statistical research in social science, where no support is sought from theory. This is also in line with much empirical work in agricultural technology, which has, in fact, much stimulated statistical research in psychology, sociology, and economics. The statistical estimation of the effects of fertilizers and weather upon crops is mostly done without the benefit of any elaborate theory of
the processes of growth and multiplication of cells.

or, simply: to estimate the length of time necessary to boil an egg, people have boiled eggs! They have not inquired too deeply into the underlying molecular processes. And Kepler "fitted" the planetary orbits before, and not after, the law of gravitation was known.

So much about the Case (c). Consider, however, case (a). The variable has not changed in the past, yet the effects of its future change have to be predicted. The "fitting" of, say (2.11) or (2.12) becomes impossible. But suppose we can estimate the parameters \(a, \beta, \delta\) of the demand function—say from data on \(y, p\) and \(x\). We can then insert \(a, \beta, \delta\) into (2.6) and (2.7), and the problem is solved.

In Case (a) we have, thus, used: (1) data on \(p, x, \) and \(y\) in addition to those on \(\delta\) and \(\tau\); (2) a "model" or "theory" consisting of equations (2.1), (1.2), (1.3), (2.4).

Note that one might also include, as a further bit of information belonging to the "model," the inequality (condition for maximum)

\[
(2.13) \quad \frac{\partial^2 z}{\partial p^2} \bigg|_{p = \hat{p}} < 0,
\]

which can be used as a "check" upon the estimates, viz., as a check upon \(\beta\), which by (2.13) must be positive. Or one might say that \(\beta > 0\) on grounds of "common sense," "introspection," "previous history."

The "model" (2.1), (1.2), (1.3), (2.4), (2.13) is also called "a priori information" in the sense that it is derived from sources other than the particular data in question (viz., the time series \(p, x, y, \delta\)).

The parameters \(a, \beta, \delta, \gamma\) of the model are called "structural parameters." When all structural parameters are known, we say that not only the model but also the "structure" is known. Thus a model is a class of structures.

We have shown that the estimation of structural parameters, or structural estimation, may help to predict the effect of changes of certain variables when the direct estimation of such effects from past data is not feasible.

**Exercise 2.1** What is the relation between the coefficients \(\beta, \delta\) in equation (2.1) and the following economic concepts: 1) elasticity of demand; 2) price-elasticity of demand; 3) income-elasticity of demand; 4) marginal propensity to consume.

**Exercise 2.2** Comment upon the fact that in (1.11) \(\theta^*\) depends on \(a, \beta\) only; while in (2.8), \(\theta^*\) depends on \(\delta\) only.

**Exercise 2.3** What are the logical similarities or differences between the partial derivative in (2.6), and the concept of "multiplier" used in economics?

**Exercise 2.4** Give examples of conditions under which equation (2.10) would be used in preference to (2.8), (2.9).
Lecture 1.

As suggested at the end of the preceding lecture, we now proceed to estimate the parameters \( a, \beta, \delta \) of the model (2.2), (2.3) from data collected through a period during which \( a, \beta, \delta \) did not change. The ultimate purpose is to predict a future value of \( T \) for changed tax-rates \( \theta, \tau \), and income \( y \). We further assume that while \( y \) did change during the period of observations, the tax-rates \( \theta \) and \( \tau \) did not. (If they had, we could predict \( T \) for given \( \theta, \tau \), \( y \) directly from past observations on these variables without estimating \( a, \beta, \delta \); this was shown in the preceding lecture).

To simplify the algebra, we shall treat only the case when the constant values of both \( \tau \) and \( \theta \) during the observation period were zero, i.e., these taxes are going to be introduced for the first time. (2.2), (2.3) become:

\[
\begin{align*}
(4.1) \quad & p = a/2\beta + (\delta/2\beta) \cdot y \\
(4.2) \quad & x = a/2 + (\delta/2) \cdot y.
\end{align*}
\]

(4.1), (4.2) constitute a model, or a "maintained hypothesis." Past observations on \( p, y, x \) will serve to find which of the particular "considered hypotheses," or structures should be accepted. E.g., consider the following systems

\[
\begin{align*}
\begin{cases}
  p = 1 + 2y \\
  x = 3 + 6y
\end{cases}
\quad ;
\begin{cases}
  p = 2 + y \\
  x = 4 + 2y
\end{cases}, \quad \text{etc., etc.,}
\end{align*}
\]

they have the following values for the three structural parameters

\( a = 6, \quad \delta = 12, \quad \beta = 3; \quad a = 8, \quad \delta = 4, \quad \beta = 2; \quad \text{etc., etc.} \)

All these systems are compatible with the "maintained hypothesis" but not all of them (perhaps only one, or even none) is compatible with both the maintained hypothesis and the data. Each of these systems can be a "considered hypothesis," viz., a special case of the "maintained hypothesis." Some of them will be rejected, others (possibly one or none) accepted in the light of data.

If (4.1), (4.2) were seriously used as a maintained hypothesis, one observation would suffice to determine \( \beta \), since \( \beta = x/p \) in any year, say, \( x_1/p_1 = \beta \). Should the corresponding ratio in a second year, \( x_2/p_2 \) be different, the model would be disproved at once. To find \( a, \delta \), one needs the observations on \( x \) and \( y \)--say \( (x_1, y_1), (x_2, y_2) \). Then, (4.2) gives two equations

\[
(4.3) \quad x_1 = a/2 + (\delta/2) \cdot y_1, \quad x_2 = a/2 + (\delta/2) \cdot y_2,
\]

from which one finds the two unknowns \( a, \delta \). If the model (4.1), (4.2) is exactly true, a third observation would not provide additional knowledge. Nor would it be of any further help to have a second observation on \( p \) and to fit (4.1) separately to the data \( (p_1, y_1), (p_2, y_2) \), thus obtaining \( a/2\beta, \delta/2\beta \) as a check on the previously found \( \beta, a, \delta \).
This exactness of the number of necessary observations is a peculiarity of "non-stochastic models" such as (4.1), (4.2). Such models are hardly realistic but can serve, to show certain "pre-statistical" problems of econometric research.

With the values for the "structural parameters" \( \alpha, \beta, \delta \) thus obtained, one can predict the results (e.g., the effect upon the tax revenue \( T \)) of given "structural changes," i.e., changes in \( \alpha, \beta, \delta \). For example, the introduction of an income tax \( \tau \) means replacing the term \( \delta y \) of the demand equation (2.1) by \( \delta (1 - \tau) y \); i.e., \( \delta \) is multiplied by a known constant, viz., by \( (1 - \tau) \). Similarly, one might ask the effect of a given (intended or expected) increase in \( \beta \), the price-slope of the demand curve, due, for example, to a (induced or spontaneous) change of tastes.

We can distinguish between controlled and uncontrolled changes of variables and call the latter "policy changes." We can also distinguish between prospective "routine changes"—i.e., changes of variables which had been changing in the past—and "structural changes," i.e., changes of variables which had been constant previously. In our model, the variable \( y \) is subject to uncontrolled routine changes. The variables \( \tau, \theta \) are subject to controlled, or policy changes, which can be either routine changes or structural changes. In order to predict the effects of given structural changes (e.g., structural policies) we need the knowledge of previous structure. But if the changes (and, in particular, policies) are of the routine type, the knowledge of structure is not needed, as effects of future changes can be predicted on the basis of experienced effects of past changes.

Reading assignment: Buffalo lectures pp. 15-20, 42-45.

Problem 4.1. Use data for a small number of years from the Statistical Abstract for U.S. to construct a numerical example which would illustrate equations (4.1), (4.2) without contradicting the facts too absurdly. Motor fuel oil (gasoline) may be an appropriate commodity for illustrative purposes.
The empirical determination of economic structural relations goes back to the late seventeenth century when Gregory King estimated the demand curve for wheat. It was a "non-parametric" estimate: King measured the relative price rise at five levels decline in the available quantity of wheat. During the nineteenth century Cournot (1838: he was the founder of mathematical economics), Jevons and Marshall urged the need for empirical measurement of economic relations, but did not contribute more than the repetition of King's estimate (which Jevons translated into a "parametric" form by fitting a hyperbola to King's five pairs of figures). Shortly before World War I, Lehel (South Africa) made a new study of the wheat demand function, and French and Italian authors made similar attempts for other commodities. After World War I, various American economists, especially Henry L. Moore, Henry Schultz, and research workers at the U. S. Department of Agriculture (M. Ezeckel, L. Bean, and others) produced a series of demand studies. The positively eloped (1) "demand curve" for steel obtained by Moore was puzzling, and by giving it the name of a "statistical demand curve" (as distinguished from the "theoretical" or "neo-classical" one) the problem was, of course, not solved. Pigou (in an Appendix to his Economics of Welfare) and Elmer Working ("What Do Statistical Demand Curves Show?," Quarterly Journal of Economics, 1927) brought more light into the problem by relating the observed values of prices and quantities of a commodity to the partial equilibrium model as discussed below. Ragnar Frisch (Pitfalls in the Statistical Study of Demand and Supply, 1933) gave a fuller account of the problem by introducing into it random fluctuations; he also formulated the "multicollinearity" problem (see below). For a fuller history of the problem up to 1938, see Henry Schultz, Theory and Measurement of Demand. The problem was for a long time confused with that of "serial correlation" of observations that succeed each other in time, and the whole discussion went under the label of "analysis of economic time series." A new impulse to clarification was given by Trygve Haavelmo in 1943, followed by the work of the Cowles Commission during 1943-47, which resulted in better understanding of the logical and mathematical nature of the problem of estimating relations from non-experimental data.

Some of the relevant ideas can be presented without introducing random variations. Such presentation is perhaps easier for economists without statistical training. Accordingly, I shall go on with "non-stochastic" models until this set of ideas is classified.

The "pitfalls" in question result from ignoring some features of the true set of relevant economic relations, or the "true model." Particularly important cases arise when

A) It is ignored that there are more variables in the studied relation than the ones taken into account.

B) It is ignored that there are more relations in the model than the one studied.

C) It is ignored that more than one structure is compatible with the same model.

We shall start with a simple case in which none of these omissions occur, and
then discuss separately the pitfalls A, B, C. Suppose the quantity supplied, \( x \), is independent of the price \( p \), e.g., supply is completely determined by rainfall (crops) or by migrations of animals (fish); and suppose that, because the commodity is perishable, the supply is equal to demand; suppose further that demand is determined by the price only:

\[
(5.1) \quad x = \alpha - \beta p,
\]

where \( \alpha, \beta \) are parameters of the demand schedule. Under these assumptions the single equation (5.1) states the model completely, with \( x \) an exogenous variable. That is, \( x \) may be determined by a relation such as

\[
(5.2) \quad x = a \text{ function of rainfall},
\]

but this relation does not involve any further variables that interest us, and may be studied separately (presumably by biologists, not economists). Graph \( V \) drawn in the \((x, p)\)-plane shows the equation (5.2) as a set of vertical lines, each corresponding to a different amount of rainfall (See Exercise 5.1). Equation (5.1), on the other hand, is represented by a single (downward-sloping) curve, since the parameters \( \alpha, \beta \) are assumed unchanged ("structural parameters"). The intersection points are labelled to number the observations in the order of their succession in time, or in space, or in some other order. Given the observations, the demand curve is obtained by connecting them (two observations suffice).

**Pitfall A.** Suppose now that the true demand equation is not (5.1) but

\[
(5.1A) \quad x = \alpha - \beta p + \delta y,
\]

where \( y \) = national disposable income; while (5.2) is, as before, the supply equation. On Graph \( V_A \) the demand function is a family of curves relating \( x \) and \( p \) (a family of "shifting" Marshallian demand curves), one curve for each observed value of \( y \). The intersection points are labelled as before. Their configuration in the \((x, p)\)-plane cannot be used to reconstruct the demand function. The latter is now a surface—specifically, a plane—not a line. But the configuration of the intersection points in the \((x, p, y)\)-space would indeed lie on the demand surface, and three such points would suffice to determine the three unknown parameters \( \alpha, \beta, \delta \), from the three equations

\[
\begin{align*}
\alpha_1 &= \alpha - \beta p_1 + \delta y_1 \\
\alpha_2 &= \alpha - \beta p_2 + \delta y_2 \\
\alpha_3 &= \alpha - \beta p_3 + \delta y_3,
\end{align*}
\]

where \((x_i, p_i, y_i), i = 1, 2, 3\), are observations made at three points of time (or, say, in three counties).

The case of Moore's upward-sloping "statistical demand curve" for steel might occur if the supply curve did not shift through the observation period, while income \( y \) did (Graph \( V_A' \)); or, more generally, if the supply curve did shift (changes in the technology, wage-rates, etc.), but less violently than the changes in the Marshallian demand curve shifted under impact of (so-called cyclical) fluctuations of national income: Graph \( V_A'' \).

**Pitfall B.** Consider the example of Lectures 2-4. There the demand function is as in (5.1.A); but the second equation of the model does not explain supply by weather. It shows, instead, how the monopolist adjusts the price, or the supply,
to the changes of national income $y$. One and only one of the equations (5.1) or (5.2) can then be used together with (5.1.4) to complete the model. (If we use both equations (5.1) and (5.2), one of them will turn out to be redundant.) We can also use, instead, a particular simple relation that results from combining (5.1), (5.2), viz., $x = \beta p$. Consider, then, the system

$$
\begin{align*}
5.1.5 & \quad x = \alpha - \beta p + \delta y \\
5.2.1 & \quad x = \beta p
\end{align*}
$$

If we ignore (5.2.1) and try to draw a surface (plane) through three observation points as suggested above, we shall fail. These points are collinear: they lie on a straight line, viz., the intersection line between the planes (5.1.5) and (5.2.1) (the latter happens to be a plane through the y-axis). Algebraically speaking, only two equations of the system (5.5) in the unknowns $\alpha$, $\beta$, $\delta$ are independent.

In general the set of points obeying a relation has $(m-1)$ dimensions less than the set of points obeying only one relation. Hence, the general term "multicollinearity." Because of the presence of random disturbances in the data, the points may show an apparently higher dimensionality than that of the true model, and the researcher may not notice, for example, that the points lie almost on a line, and proceed to "fit" a plane: hence, multicollinearity is a "pitfall."

**Pitfall C.** In the case just discussed, Pitfall C is not present. There is only one set of values for $\alpha$, $\beta$, $\delta$ that is consistent with the data and with the model (5.1.5), (5.2.1). True, we have just shown that these values (the "structure") cannot be obtained by using the singular relation (5.1.5) only; but they can be obtained if we take account of both relations. This was shown in Lecture 1. But consider, instead, the following model of supply-and-demand in a competitive market:

$$
\begin{align*}
5.3.0 & \quad x = \alpha_d - \beta dp \\
5.2.0 & \quad x = \alpha_s + \beta sp
\end{align*}
$$

where the parameters $\alpha_d$, $\beta_d$ describe the structure of demand (behavior of buyers), and $\alpha_s$, $\beta_s$ describe the structure of supply (behavior of suppliers). Multicollinearity (Pitfall C) is present in this case: the observations lie in a single point (dimensionality zero), while the relation we want to determine—say, the demand curve—is a line (dimensionality one). In addition, the model is such that an infinite number of "structures," i.e., of values of the $\alpha$'s and $\beta$'s is compatible with the model and the data: there is an infinite number of pairs of straight lines through a given point of observations. Algebraically, any straight line that goes through the same point as the two lines (5.1.0), (5.2.0) can be regarded as a "weighted average" between these two lines. (The weights need not be all positive.) E.g., if a certain pair of values for $x$, $p$ satisfies (5.1.0), (5.2.0), then it will also satisfy the equation formed by multiplying (5.1.0) by 1/3, multiplying (5.2.0) by 2/3, and adding

$$
\begin{align*}
90 = [(1/3)\alpha_d + (2/3)\alpha_s] + [(1/3)\beta_d + 2/3(\beta_s)]
\end{align*}
$$

See Graph V C, where the demand and supply curves are drawn as solid lines, while the intercept and slope of the dotted line are weighted averages of the intercepts and slopes of the demand and supply curves.)
The model (5.1.C), (5.2.C), together with the data, is not sufficient to identify the particular structure that has generated the observed data. Consider now the model

\[
\begin{align*}
(5.1.C') \quad x &= \alpha_d - \beta_d y + \delta_d y \\
(5.2.C') \quad x &= \alpha_s + \beta_s p
\end{align*}
\]

and apply again the test of "weighted averages" (or "weighted sums" which amounts to the same thing). Generate a new pair of relations by weighting the original system twice: first multiply the two equations by \(m_1, m_2\) (not both zero), respectively; and then by \(n_1, n_2\) (not both zero), respectively. We obtain

\[
\begin{align*}
(5.0''') \quad \begin{cases} 
(m_1 + m_2)x &= (m_1 \alpha_d + m_2 \alpha_t) + (-m_1 \beta_d + m_2 \beta_s)p + m_1 \delta_d y \\
(m_1 + n_2)x &= (m_1 \alpha_d + n_2 \alpha_s) + (-m_1 \beta_d + n_2 \beta_s)p + n_1 \delta_d y
\end{cases}
\end{align*}
\]

Upon dividing by the coefficient of \(x\), each of the new equations is indistinguishable from (5.1.C'). Therefore, the demand plane cannot be identified by the model and data. But the supply relation can: it is distinguished from any other relation compatible with the model and the data (and generated by giving appropriate values to \(m_1, m_2, n_1, n_2\)). The supply relation is identified by the fact that it is the coefficient of \(y\) is zero, while in both equations of (5.0'') \(y\) occurs with a non-zero coefficient.

Geometrically: the observation points compatible with model (5.1.C'), (5.2.C) lie on a straight line (i.e., Pitfall II would trap any research worker trying to fit the demand plane directly to observations). There is an infinite number of planes through this straight line. But only one of these planes is parallel to the \(y\)-axis. This is the supply plane.

Similarly, we can see that both equations of the model (5.1.B), (5.2.B) are identifiable. Weighted sums of these two equations are:

\[
\begin{align*}
(5.0'') \quad \begin{cases} 
(m_1 + n_2)x &= m_1 \alpha + (n_2 - m_2) \beta p + m_1 \delta y \\
(n_1 + n_2)x &= m_1 \alpha + (n_2 - n_1) \beta p + n_1 \delta y
\end{cases}
\end{align*}
\]

None of these new equations has the properties of (5.1.B) or (5.2.B). (5.2.B) is identified by having zero-intercept, and a zero-coefficient of \(y\) (it is a plane through the \(y\)-axis). Further, in (5.1.B) \(p\) has the same coefficient (with opposite sign) as in (5.2.B), which is not the case in (5.0''). Once the plane (5.2.B) is obtained, i.e., once \(\beta\) is determined, the plane (5.1.B) is also shown to be identifiable, since there is only one plane that goes through a certain line and that has one of its slopes determined (viz., \(-\beta\)).

We have thus studied cases in which all, one, or none of the equations constituting a structure was identifiable. There may also be cases that only some of the structural parameters of certain equations can be identified.

**Exercise 5.1.** Supply all the graphs referred to in the text.

**Exercise 5.2.** The system (5.1.B), (5.2.B) is only one of possible ways of writing the model of Lecture 4. What are the other ways? For each of the systems that render that model, construct graphs showing each of the two equations in each of the three planes (\(x, p\), \(x, y\), \(p, y\)), using the numerical values obtained in Exercise 5.1.
Exercise 5.3. Apply the weighted sums test to check, for each of the systems discussed in Exercise 5.2, the identifiability of the equations.

Exercise 5.4. A. C. Pigou has suggested the following method to obtain the Marshallian demand curve: in the \((P, x)\) plane, draw through three points of observation that correspond to three consecutive years, three equidistant parallel lines; they will represent the Marshallian demand curves. What are Pigou's assumptions?
Lecture 7. See Buffalo Lectures, pp. 49-52.

Exercise 7.1. Is the following "monopoly model" overidentified:

\[ x = c + \beta p + \delta y \]

\[ x = \beta p \]

Lecture 8. See Buffalo Lectures, pp. 45-49. As a summary, the following table answers the question: "When is structural estimation from past data necessary?" and thus answers the question "Why has the identification problem arisen in economics and was generally ignored in experimental science?"

The binary system is used to classify the following cases:

0: No structural changes are expected or intended. Therefore, structural estimation unnecessary.

1: Structural changes are expected or intended: see 10, 11.

10: Specific experiments are possible; they permit the estimation of the relevant equation of the reduced form. Therefore, structural estimation unnecessary.

11: Specific experiments are impossible: see 110, 111.

110: General experiments are possible; they permit the estimation of the relevant equation of the structure. Therefore, structural estimation from past data unnecessary.

111: General experiments impossible. Therefore, structural estimation from past data necessary.

1110: The structure is (or at least the relevant structural parameters are) identifiable. Therefore, structural estimation from past data necessary and possible.

1111: The relevant structural parameters are not identifiable. Therefore, structural estimation from past data necessary but impossible.

Exercise 8.1. Consider a monopolistic market without sales tax and income tax (as in Exercise 7.1), and suppose one has to estimate the effects of introducing a sales tax for the first time. Use this case to exemplify the cases 0, 10, 110, 1110 above, describing hypothetical experiments in the cases 10, 110.
Lectures 9 and 10: Discussion of problems from previous lectures.

Lecture 11. Vector notation. Endogenous variables \( y = (y_1, \ldots, y_G) \). Exogenous variables \( z = (z_1, \ldots, z_H) = (z^c, z^u) \), where \( z^c \) and \( z^u \) denote respectively the controlled and the uncontrolled sets of exogenous variables.

Parameters of the \( g \)-th structural equations: \( a(g) = (a_{g1}, a_{g2}, \ldots) \). All structural parameters: \( a = (a(1), \ldots, a(G)) = (a^c, a^d) \), where \( a^c, a^d \) denote respectively, the controlled and the uncontrolled subsets of \( a \). System of structural non-stochastic equations:

\[
\mathbf{S}_g(y, z; a(g)) = 0, \; g = 1, \ldots, G.
\]

Its reduced form is:

\[
y_g = \mathcal{Y}_g(z; \Pi(g)), \; g = 1, \ldots, G,
\]

where \( \Pi(g) \) is the set of parameters relating the \( g \)-th endogenous variable to all exogenous variables. We write \( \Pi = (\Pi(1), \ldots, \Pi(G)) \).

A diagram was given (reproduced in Exercise 13.2) to show when and why structural estimation is needed. We want to predict \( y \) from \( z \) because we want to choose such values of the controlled \( z^c \) as to maximize the utility (welfare) \( U(y) \) a function of \( z \). (As a part-problem we are often interested in the effects of changes in \( z \)—say \( \delta y/\delta z \); the so-called "multipliers." ) If structure \( c \) is known to remain unchanged, i.e., if all expected changes in \( y \) will be due to changes in variables that were changing during the observation period (namely, \( z \)), then the knowledge of \( c \) is not necessary. The observed values of \( y, z \) always permit the finding of \( \Pi \) (the reduced form is always identifiable), and this permits the predicting of \( y \) from \( z \). Suppose, however, we know that the structural parameters \( a \) (quantities which did not change during the observation period) are going to change in a known way, viz., into

\[
a^* = \mathcal{T}(a);
\]

(\( \mathcal{T} \) is a known transformation). Then \( y \) will depend on \( z \) in a new way, as described by the new reduced form \( \Pi^* \). This, in turn, depends on the new structure, \( a^* \). Since we know \( \mathcal{T} \), we can determine \( a^* \) (and therefore \( \Pi^* \)) if we know \( a \). Thus, we need structural estimation whenever we expect or intend the structure to change in some known way.
Lecture 12. A random variable takes its different values with certain probabilities. If \( y \) takes discrete values \( \eta', \eta'' \) ... with probabilities \( p', p'' \) ..., respectively, the set of pairs \((\eta', p'), (\eta'', p'') \) defines a probability distribution function, \( p = f(y) \). We shall use, in general, Roman letters for random variables, and Greek letters for other variables and for constants (a rule which we have already broken by using Roman \( p \) above).

For a continuous random variable \( y \), we first define the cumulative distribution function, viz., the probability that \( y \) exceeds \( \eta \), \( \Pr(y > \eta) = F(\eta) \). Then \( \Delta F = F(\eta + \Delta\eta) - F(\eta) \) is the probability that \( y \) lies in the "bracket \( \eta \) to \( \eta + \Delta\eta \)." Let now the bracket become "thinner," i.e., let \( \Delta\eta \) diminish. If \( \Delta F/\Delta\eta \) tends to a limit, the function \( F \) is called a differentiable one, and the limit itself a derivative, \( df/d\eta = f(\eta) \), say. We call the derivative of the cumulative probability function, the probability density function. [Examples of cumulative frequency and of frequency density were given, using income distribution.]

We shall write \( f(y) \) to denote probability if \( y \) is discrete, or probability density if \( y \) is continuous. Either will be called the "distribution of \( y \)."

Using the vector notations of Lecture 12, the following distribution functions were defined:

1) univariate non-conditional distribution \( f(y|_{g}) \)
2) " conditional " \( f(y|_{g}, z) \)
3) joint non-conditional " \( f(y) \)
4) " conditional " \( f(y|_{z}) \).

Of these, 1) was defined above; 2) can be exemplified, for \( y_{g} \) discrete, as follows:

<table>
<thead>
<tr>
<th>Value of ( y ) has the value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \eta' ) ( \eta'' ) ( \eta''' ) ...</td>
</tr>
<tr>
<td>when ( z=1 ) ( p'(1) ) ( p''(1) ) ( p'''(1) ) ...</td>
</tr>
<tr>
<td>&quot; ( z=2 ) ( p'(2) ) ( p''(2) ) ( p'''(2) ) ...</td>
</tr>
<tr>
<td>etc. ( = ) ( = ) ( = ) ( = ) ( = ) ( = ) ( = )</td>
</tr>
</tbody>
</table>

where "\( z=1 \)" means "combination No. 1 of the variables \( z_{1}, z_{2}, ... \); for example, "\( z=1 \)" may mean "\( z_{1} = 0, z_{2} = 5, z_{3} = 1 \)" etc.

Examples of joint distributions are given below in (12.2) (non-conditional) and (12.6) (conditional). The last case is the most general one, and of great importance for economists. Joint distribution is also called multivariate.
In stochastic economics, one or more of the structural equations contain random disturbances, in general not directly measurable. Each disturbance is the aggregate effect of a large number of exogenous variables, each of these variables taking small values only. The presence of disturbances in the structural equations should explain the fact that endogenous variables themselves behave like random variables. The distribution of the observables depends on 1) the distribution of the disturbances and 2) the structural equations. In the following, \( x \) and \( y \) are endogenous variables, \( z \) is an exogenous variable, \( u_1 \) and \( u_2 \) are disturbances, and \( \alpha, \beta, \gamma \) are structural parameters:

\[
\begin{align*}
\text{Supply} & \quad x = \gamma y + u_1 \\
\text{Demand} & \quad x = -\beta y + \gamma z + u_2
\end{align*}
\]

where \( x \) = quantity per head, \( y \) = price, \( z \) = income per head.

Distribution of disturbances:

\[
\begin{array}{ccc|c}
 & u_1 & +1 & -1 & Pr(u_1) \\
\hline
u_2 & 1/6 & 2/6 & 3/6 & 5/6 \\
\hline
\hline
\text{Pr}(u_2) & 4/6 & 2/6 & 1 & 1
\end{array}
\]

(Note that \( u_1, u_2 \) are "not independently distributed": the distribution of \( u_1 \) is not the same when \( u_2 = +1 \) and when \( u_2 = -1 \).)

The above equations (12.1) (with \( \alpha, \beta, \gamma \) having certain numerical values) plus the distribution (12.2) of the \( u \)'s constitute stochastic structure. The reduced form is also stochastic (solving the two structural equations for \( y, x \)):

\[
\begin{align*}
\pi_2 &= \frac{\gamma}{\alpha + \beta} \\
\pi_1 &= \frac{\gamma}{\alpha + \beta} \\
\end{align*}
\]

Thus, given \( z \), the joint distribution of \((y, x)\) is easily derived from that of \((v_2, v_1)\). Suppose, for example, \( \alpha = \beta = \frac{1}{2}, \gamma = 1. \) Then

\[
\begin{align*}
\pi'_2 &= 1, & v_2 &= u_2 - u_1 \\
\pi'_1 &= 1/2, & v_1 &= (u_2 + u_1)/2
\end{align*}
\]

First derive the joint distribution of \((v_1, v_2)\) from that of the \( u \)'s. (12.2) gives the probabilities of three alternative events:
\[ u_1 = +1, \ u_2 = +1 ; \ \text{then} \ v_1 = 1, \ v_2 = 0; \ \text{the probability is} \ 3/6; \]
\[ u_1 = +1, \ u_2 = -1 ; \ " \ v_1 = 0, \ v_2 = 1; \ " \ " \ " \ 1/6; \]
\[ u_1 = -1, \ u_2 = +1 ; \ " \ v_1 = 0, \ v_2 = -1; \ " \ " \ " \ 2/6. \]

Therefore the joint distribution of the \( v \)'s is

\[
\begin{array}{ccc}
\text{Pr}(v_2) \\
\text{v_1} & 0 & 1 \\
\text{v_2} & \\
-1 & 2/6 & 0 & 2/6 \\
0 & 0 & 3/6 & 3/6 \\
+1 & 1/6 & 0 & 1/6 \\
\text{Pr}(v_1) & 3/6 & 3/6 & 1 \\
\end{array}
\]

We can now derive from (12.3), (12.4), (12.5) the joint distribution of \( x, y \) when \( z \) is, say, 10; or \( z = 20; \)

\[
\begin{array}{ccc}
z = 10 & & z = 20 \\
\text{y} & & \\
\text{x} & 5 & 6 & 10 & 11 \\
9 & 2/6 & 0 & 19 & 2/6 & 0 \\
10 & 0 & 3/6 & 20 & 0 & 3/6 \\
11 & 1/6 & 0 & 21 & 1/6 & 0 \\
\end{array}
\]

Suppose we have numerous observations on \( x, y \), made at various levels of \( z \). Then the observed frequencies of the various combinations of \( (x, y) \) for a given \( z \) will tend to be in proportion to the joint probabilities listed in (12.6). A "scatter diagram" in the \( (x, y) \) plane, constructed for \( z = 10 \) will consist of three clusters of points: half of all points will lie around \( (x = 6, y = 10) \); one-third around \( (x = 5, y = 9) \); and one-sixth around \( (x = 5, y = 11) \). And similarly for other values of \( z \).  

As long as the structure is not expected to change, the effect of \( z \) upon \( x \) and \( y \) can be predicted from the observed frequency distribution (12.6), which (with a large sample) tends to be \( f(x, y | z) / f(z) \). For example, there are rules of thumb as to the effect of national income upon the demand for (and possibly price for) meat—corresponding to the reduced form (12.3) plus the joint distribution of \( v \)'s (12.5).  

However, observed frequency distributions cannot serve for prediction if the structure changes, i.e., if either \( c, \ \beta, \ \gamma \) changes or if the distribution of the disturbances changes. As an example for the latter: \( u_1 \) may become more
stable if a hardier plant variety is invented; \( u_2 \) may become less stable if a substitute for the product is invented. If known structural changes are expected or intended, and the old structure is also known, then the new structure can be obtained, and consequently—by computations similar to those above—the new distribution can be obtained for prediction purposes.

The problem is thus to obtain knowledge of structure from the knowledge of the distribution of the observables. The latter can always be estimated from the frequency distribution of the observables; but the step from \( f(y \mid z) \) to the knowledge of \( \sigma, \beta, \gamma \) and (12.2) involves the problem of identification (analogous to that of the non-stochastic cases) as well as other problems.
Exercise 13.1. The demand for a product is \( x = -\beta p + \delta y - \nu \), where \( p \) = price, \( y \) = income, and \( u \) = a random term with the following distribution: \( \Pr(u = 2) = \Pr(u = -2) = .5 \). The product is made and sold by a monopolist at constant total cost. When setting the price to maximize profit, he always wrongly assumes demand to be \( -\beta p + \delta y + \nu \), where \( \nu \) is independent of \( u \) and has the following distribution: \( \Pr(v = 1) = \Pr(v = -1) = .5 \). Suppose \( \delta = 2 \), \( \beta = 0.5 \) and derive the following distributions:

- Distribution of \( p \) when 1) \( y = 10 \) ; 2) \( y = 20 \)
- Distribution of \( x \) when 1) \( y = 10 \) ; 2) \( y = 20 \)
- Joint distribution of \( p \) and \( x \) when 1) \( y = 10 \) ; 2) \( y = 20 \).

Exercise 13.2. In Lecture 11, the following diagram reproduced below showed the need for structural estimation assuming non-stochastic economics. Modify the notations so as to make the diagram applicable to stochastic economics, and explain.

```
   \( \mathcal{J} \)
   \( a \) --\( \rightarrow \) \( a_s = \mathcal{J}(a) \)

Observed
\( y, z \) \( \rightarrow \) \( \Pi \)
\( \rightarrow \) \( \mathcal{J}_z \)
\| Predicting \y from \z Predicting
\| \y from \z and \mathcal{J}
```
The problems of lecture 13 were discussed. In the formulation of Exercise 13.1, one possible ambiguity must be removed. In the problem, the behavior of the monopolist is stochastic: the intercept \( v \) of his estimate of the demand function is a random variable; but in his eyes, the demand function for a given year is not a stochastic one; through market research (consumers' surveys) he tries to establish the exact current demand function, and error by the amount \( u - v \). Another interpretation offered was that of a monopolist who considers the demand function for a given year as a stochastic one. He believes to know the distribution of \( v \). In this case his profit is also stochastic. He may then try to maximize the expected (in the statistical sense) value of profit; or to maximize the actual value of the utility of profit; or to maximize the actual value of the profit net of progressive income-tax. (In the last two cases, high profits will be given smaller weights than losses or low profits.)

As to Exercise 13.2, the required diagram for the stochastic case differs from that in the non-stochastic as follows. Let \( Pr \) mean "joint probability of a (random) vector." Replace \( u \) by \( u_s \) and \( Pr(u_s) \). Replace \( c_0 \) by \( c_0 \), and \( Pr(u_s) \). Replace \( \Pi \) by \( \{old \Pr(y | x)\}, \) or by \( \Pi \) and \( \Pr(v) \), where \( v \) denotes the random residuals in the old reduced form. Replace \( \Pi_0 \) by \( \text{New } Pr(y | x) \) or by \( \Pi_{01} \) and \( \Pr(v) \), where \( v \) denotes the random residuals in the new reduced form.

In Lecture 12 and Exercise 11.1, non-parametric forms of distributions were used. We now proceed to parametric form and discuss familiar distribution parameters.

1) non-conditional univariate distribution, \( f(y) \) where \( Pr(\gamma_i) = \pi_i \).

Parameters: Mean (= expectation) = \( \Sigma \gamma_i \pi_i = \mathcal{E} y \)
Median. Mode.
Variance (second moment about the mean) \( \sigma^2_y = \sigma_{yy} = \mathcal{E} (y - \mathcal{E} y)^2 \)
Standard deviation = \( \sqrt{\sigma_{yy}} \)
Mean deviation. Range.
Third moment about the mean = \( \mathcal{E} (y - \mathcal{E} y)^3 \), etc.

2) non-conditional bivariate distribution \( f(y_1, y_2) \)
where \( Pr(\gamma_1^1) = \pi_1^1, \ Pr(\gamma_2^2) = \pi_2^2, \ Pr(\gamma_1^1, \gamma_2^2) = \pi_{12} \);
\( \mathcal{E} y_g = \Sigma \gamma_i \pi_i, \ g = 1, 2. \)
\( \sigma^2_{y_1 y_2} = \sigma^2_{y_1 y_2} = \mathcal{E} (y_1 - \mathcal{E} y_1)(y_2 - \mathcal{E} y_2) \)
Covariance = \( \sigma_{y_1 y_2} = \sigma_{12} = \mathcal{E} (y_1 - \mathcal{E} y_1)(y_2 - \mathcal{E} y_2) \).
Higher cross-moments, for example: \( \mathcal{E} (y_1 - \mathcal{E} y_1)^2 (y_2 - \mathcal{E} y_2) \)
Correlation coefficient = \( \rho = \sigma_{12}/\sqrt{\sigma_{11} \sigma_{22}} \).
2a) Extension to any number G of variates:

G expectations \( \mathcal{E} y_g, g = 1, \ldots, G \)

G variances and covariances \( \sigma_{gh}, g, h = 1, \ldots, G \):

\[
\begin{pmatrix}
\sigma_{11} & \sigma_{12} & \cdots & \sigma_{1G} \\
\sigma_{21} & \sigma_{22} & \cdots & \sigma_{2G} \\
\vdots & \vdots & \ddots & \vdots \\
\sigma_{G1} & \sigma_{G2} & \cdots & \sigma_{GG}
\end{pmatrix} = \text{covariance matrix (symmetrical, since } \sigma_{gh} = \sigma_{hg})
\]

Higher moments. (Also: use of medians, ranges, etc.)

3) Conditional uni-variate distribution, \( f(y \mid z) \) where \( y \) is a random variable, but \( z \) is a non-random variable. The distribution parameters of \( y \) are functions of \( z \). In particular,

\( \mathcal{E} y \mid z = R(z), \text{ say: the regression equation of } y \text{ on } z. \)

If \( R(z) \) is linear, \( \mathcal{E} y \mid z = \pi_0 + \pi_1 z \), \( \pi_1 \) is called regression coefficient of \( y \) on \( z \). One can also discuss, for example, \( \sigma_{yy} \mid z = V(z), \text{ say, etc.} \)

Moreover we can generalize to the case when \( z \) is a vector, \( z = (z_1, \ldots, z_K) \), while \( y \) is still a scalar. If the regression equation is linear

\[ \mathcal{E} y \mid z = R(z) = \pi_0 + \pi_1 z_1 + \cdots + \pi_K z_K, \]

the \( \pi \)'s are called partial regression coefficients.

4) Conditional multivariate distribution——

\[ f(y_1, \ldots, y_G \mid y_{H+1}, \ldots, y_G, z_1, \ldots, z_K). \]

This is the most general case. The distribution of one or more random variables is regarded as depending on the values taken by one or more of the remaining random variables and by the non-random variables. The latter may or may not be present. Example: \( H = 2, G = 4, K = 1 \). Parameters of \( f(y_1, y_2, y_3, y_4, z_1) \) are

\[ \mathcal{E} y_1 \mid y_3, y_4, z_1 = R(y_3, y_4, z_1); \text{ regression function (possibly linear) of } y_1 \text{ on } y_3, y_4, z_1. \]

Conditional variance: \( \sigma_{y_1y_1} \mid y_3, y_4, z_1; \) conditional correlation coefficient:

\[ \rho_{y_1y_2} \mid y_3, y_4, z_1; \]

Consider \( f(y_1, y_2) \) and suppose \( \mathcal{E} y_2 \mid y_1 = R(y_1) = \pi_{20} + \pi_{21} y_1 \); then \( \pi_{21} \) is regression coefficient of \( y_2 \) on \( y_1 \); \( \pi_{12} \) is similarly defined.
Exercise 114.1. Consider the bivariate distribution

| \( y_1 \) | \(-1\) | 0 | \(+1\) | \( \Pr(y_2) \) | \( \mathbb{E}y_1|y_2 \) | \( \sigma^2_{y_1} | y_2 \) |
|-----|-----|-----|-----|-----|-----|-----|
| \(-1\) | 0 | \(\frac{1}{8}\) | \(\frac{1}{8}\) | ? | ? | ? |
| \(+1\) | \(\frac{1}{8}\) | \(\frac{1}{8}\) | \(\frac{1}{2}\) | ? | ? | ? |

| \( \Pr(y_2) \) | ? | ? | ? | 1 | \( \mathbb{E}y_2 = ? \) | \( \sigma^2_{y_2} = ? \) |
| \( \mathbb{E}y_2|y_1 \) | ? | ? | ? | ? | ? |
| \( \sigma^2_{y_2} | y_1 \) | ? | ? | ? | ? | ? |

Replace question marks by numbers. Plot both regression equations. Plot the two equations expressing the variance of one of the variables as a function of the value of the other.

Exercises 114.2 and 114.3. Same question as in 114.1, for the following three distributions:

<table>
<thead>
<tr>
<th>( y_1 )</th>
<th>(-1)</th>
<th>?</th>
<th>(+2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>114.2:</td>
<td>3</td>
<td>(\frac{1}{3})</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>(\frac{2}{3})</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( y_1 )</th>
<th>(-1)</th>
<th>?</th>
<th>(+2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>114.3:</td>
<td>1</td>
<td>(\frac{1}{3})</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>(\frac{2}{3})</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( y_1 )</th>
<th>(-1)</th>
<th>?</th>
<th>(+2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>114.4:</td>
<td>3</td>
<td>(\frac{1}{3})</td>
<td>(\frac{2}{3})</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>(\frac{1}{3})</td>
<td></td>
</tr>
</tbody>
</table>

Exercise 114.5. In which of the above four distributions have you found: 1) zero correlation; 2) perfect correlation; 3) stochastic independence between \( y_1 \) and \( y_2 \); 4) stochastic independence between \( y_2 \) and \( y_1 \); 5) non-linear functional dependence; 6) linear functional dependence; 7) one of the regression functions a constant; 8) both regression functions constants; 9) one of the conditional variances constant; 10) both variances constant.
Lecture 15.

The concepts exemplified in Exercise 15.5 were discussed. In particular, using notations of Exercise 1.1,

\[
\begin{cases}
E[y_1y_2|z] = 0 \implies (y_1, y_2 \text{ statistically independent}) \implies (E[y_1|y_2=\text{const}] = (E[y_2]=0) \\
E[y_1y_2] = 1 \implies (y_1, y_2 \text{ linearly dependent}) \implies (y_1, y_2 \text{ functionally dependent}).
\end{cases}
\]

Consider the regression equation of a single random variable \(y\) upon \(k\) non-random ones, e.g. (taking the simplest case \(k = 1\), and writing \(z_1 = z\))

15. \[E[y|z] = \pi_0 + \pi z.\]

Denote by \(v\) the random residual

\[v = y - E[y|z] = y - (\pi_0 + \pi z).\]

Then the regression constants \(\pi_0, \pi\) and the residual \(v\) derived therefrom have the following properties:

15.3 \[E(v|z) = 0, \text{ since for any value of } z, (15.2) \implies E[v = E[y|z] - E[y|z];\]

15.4 \[E(vz|z) = 0, \text{ since } E[vz = zE[v|z]. \text{ That is, } v \text{ and } z \text{ are non-correlated.}\]

Furthermore, let \(z\) take \(k\) values \(z^{(1)}, \ldots, z^{(N)}\). Then

\[E \sum_{i=1}^N v^2z^{(n)} \leq E \sum_{n=1}^N v^2z^{(n)}, \text{ that is: the regression constants } \pi_0, \pi;\]

make the sum of squares of residuals a minimum. Proof:

\[v = y - (\pi_0 + \pi z); \text{ by (15.3)}\]

\[0 = E[v|z^{(n)} = E[y|z^{(n)} - \pi_0 - \pi z^{(n)}; \text{ and by (15.4)}\]

\[0 = E[vz|z^{(n)} = E[yz|z^{(n)} - \pi_0 z^{(n)} - \pi[z^{(n)}]z. \text{ Now sum each of the last two equations over } N \text{ values of } z^{(n)} \text{ (thus } \Sigma \text{ means } \sum_{n=1}^N):\]

\[0 = E \sum_{n=1}^N v^2z^{(n)} - N\pi_0 - \pi \Sigma z^{(n)}\]

or \[0 = E \sum_{n=1}^N v^2z^{(n)} - \pi_0 \Sigma z^{(n)} - \pi \Sigma[z^{(n)}]^2.\]

But in these equations one recognizes the "normal equations" obtained by minimizing \(E \sum v^2z^{(n)}\) with respect to \(\pi_0\) and \(\pi\), viz.,
\[ \frac{\partial}{\partial z} \mathcal{E}_z v^2|_z(n) = 0 ; \quad \frac{\partial^2}{\partial x^2} \mathcal{E}_x v^2|_z(n) = 0. \]

This proves (15.5), the "least squares property" \( \mathcal{F}_0, \mathcal{F}_1 \). This would apply, for example, to the constants \( \mathcal{F}_1, \mathcal{F}_2 \) in (12.3), where \( z \) was independent of the random variables \( v_1, v_2 \). But it would not apply to the parameters \( a, b \) in (12.4) because \( y \) is not independent of (and, therefore, in general, correlated with) \( u_1, u_2, v_1 \) depends on \( u_1, u_2 \) as shown in (12.3). Therefore, the condition (15.4) is not satisfied. The structural parameters \( a, b \) do not have the least squares property.

Some distribution parameters of linear combinations. Suppose \( x = y_1 + y_2 \). Then
\[ \mathcal{E} x = \mathcal{E} y_1 + \mathcal{E} y_2 \]

and

\[ \mathcal{F}_{xx} = \mathcal{F}_{y_1 y_1} + \mathcal{F}_{y_2 y_2} + 2 \mathcal{F}_{y_1 y_2}. \]

This is easily generalized to the case \( x = \sum g k g y_g \); then
\[ \mathcal{E} x = \sum g k g \mathcal{E} y_g \]

and
\[ \mathcal{F}_{xx} = \sum_{g=1}^{G} \sum_{h=1}^{G} k_g k_h \mathcal{F}_{y_g y_h}. \]

Thus the variance of a linear function of random variables is a linear function of covariances (including variances). Normal distribution of \( G \) random variables is fully described by the following parameters: 1) the \( G \) expected values \( \mathcal{E} y_g, g = 1, \ldots, N \); 2) the \( G(G-1)/2 \) variances and covariances, i.e., the matrix (14.1). All other parameters can be derived from these. For example, in the case \( G = 1 \), the probability density of the single variable \( y \) is

\[ f(y) = \left(2 \pi \sigma_{yy}\right)^{-\frac{1}{2}} \exp\left(-\frac{(y - \mathcal{E} y)^2}{2 \sigma_{yy}}\right), \]

the two parameters are \( \mathcal{E} y \) and \( \sigma_{yy} \). The higher odd moments such as \( \mathcal{E} (y - \mathcal{E} y)^3 \) are all zero; the fourth moment \( \mathcal{E} (y - \mathcal{E} y)^4 = 3 \sigma_{yy}^2 \), etc.

The most useful form of writing the joint normal distribution, for \( G > 1 \) is a form that "degenerates" into (15.7) for \( G = 1 \). This requires matrix notation. More important for our purposes are the following simple properties of the normal distribution:

1) If \( y_1 \) and \( y_2 \) are each normally distributed, then \( y_1 + y_2 \) is also normally distributed. More generally, the linear combination \( x = \sum k_g y_g \) is normally distributed; its mean and variance are, as always, given by (15.9), (15.10). If \( x = y_1 + y_2 \) and the \( y \)'s are non-correlated, \( \mathcal{F}_{xx} = \mathcal{F}_{y_1 y_1} + \mathcal{F}_{y_2 y_2} \).
2) $\sigma_{12} = 0$ is not only a necessary (as in (15.1)) but also a sufficient condition for statistical independence between two jointly normally distributed variables.

3) If $Y_1, Y_2$ are jointly normally distributed, then both regression equations are linear; the regression coefficient $\hat{\beta}_{12} = \sigma_{12}/\sigma_{22}$; and $\hat{\beta}_{12}$ has the least squares property.
Exogenous and endogenous variables in stochastic economics. Of the following four equations the first two may describe "economic laws," the other two some "non-economic laws," e.g., laws of political science, meteorology, etc.:

\[
\begin{align*}
\begin{cases}
\beta_1 y_1 + \beta_2 y_2 + \gamma_{11} z_1 + \gamma_{12} z_2 = u_1 \\
\beta_3 y_1 + \beta_4 y_2 + \gamma_{21} z_1 + \gamma_{22} z_2 = u_2 \\
\delta_{11} z_1 + \delta_{12} z_2 = w_1 \\
\delta_{21} z_1 + \delta_{22} z_2 = w_2
\end{cases}
\end{align*}
\]

In the non-stochastic case, the \( u \)'s and \( w \)'s would be constants; and the fact that the \( y \)'s do not enter the last two equations would guarantee that "the \( z \)'s influence the \( y \)'s but are not influenced by them," and would justify treating the \( z \)'s as exogenous variables. In the stochastic case, the absence of \( y \)'s from the last two equations does not suffice to make them independent from \( z \) (in the here appropriate, statistical sense). Consider the reduced form:

\[
\begin{align*}
\begin{cases}
\tilde{y}_1 = \pi_{11} z_1 + \pi_{12} z_2 + v_1 \text{ (a linear combination of } u_1, u_2), \\
\tilde{y}_2 = \pi_{21} z_1 + \pi_{22} z_2 + v_2 \text{ (another linear combination of } u_1, u_2),
\end{cases}
\end{align*}
\]

\[
\begin{align*}
\begin{cases}
z_1 = \beta_{1z} \tilde{y}_1 + \beta_{2z} \tilde{y}_2 \text{ (a linear combination of } w_1, w_2), \\
z_2 = \beta_{3z} \tilde{y}_1 + \beta_{4z} \tilde{y}_2 \text{ (another linear combination of } w_1, w_2).
\end{cases}
\end{align*}
\]

Hence \( \Pr(y_1, y_2 \mid z_1, z_2) \) depends on \( \Pr(u_1, u_2) \):

\[
\Pr(z_1, z_2) = \Pr(w_1, w_2)
\]

and therefore, in general, \( \Pr(y_1, y_2) \) depends on \( \Pr(u_1, u_2 \mid w_1, w_2) \). But suppose \( (u_1, u_2) \) are statistically independent of \( (w_1, w_2) \), i.e.,

\[
\begin{align*}
\Pr(u_1, u_2 \mid w_1, w_2) = \Pr(u_1, u_2)
\end{align*}
\]

Then \( \Pr(y_1, y_2) \) depends on \( \Pr(u_1, u_2) \) but not on \( \Pr(w_1, w_2) \). In this case it does not matter whether \( w_1, w_2 \) (and consequently \( z_1, z_2 \)) are random or non-random variables: we can treat \( z_1, z_2 \) as if they were non-random.

Moreover, \( \Pr(z_1, z_2) \) depends in this case on \( \Pr(w_1, w_2) \) but not on \( \Pr(u_1, u_2) \); neither, therefore, on \( \Pr(v_1, v_2) \). Therefore
\( \sum_{12} = \begin{pmatrix} 
\sigma_{11} & \sigma_{12} & 0 & 0 \\
\sigma_{12} & \sigma_{22} & 0 & 0 \\
0 & 0 & \sigma_{11} & \sigma_{12} \\
0 & 0 & \sigma_{12} & \sigma_{21} 
\end{pmatrix} \),

where the non-primed \( \sigma \)'s are distribution parameters of \((u_1, u_2)\), and the primed \( \sigma \)'s are those of \((w_1, w_2)\). We have found that all but the upper half of the \( A \)-matrix, and all but the northwest corner of the \( \Sigma \)-matrix can be neglected in explaining the observed distribution \( f(y_1, y_2; z_1, z_2) \).
Lecture 17.

Consider the following model

\begin{align*}
\text{Demand: } & y - \alpha x - \alpha_0 = u \\
\text{Supply: } & y - \beta x - \beta_0 = v
\end{align*}

(17.1)

Distribution of $u, v$ is known to be normal, and $E u = E v = 0$.

We shall write the reduced form of (17.1) and show that the parameters of the reduced form can be determined from the (infinite) sample of $(x, y)$. The reduced form of (17.1) is

\begin{align*}
(17.2) & \quad x = \pi_x + w_x \\
& \quad y = \pi_y + w_y
\end{align*}

where

\begin{align*}
\pi_x &= (\alpha_0 - \beta_0)/(\beta - \alpha) \\
& \quad \pi_y = (u - v)/(\beta - \alpha) \\
\pi_x &= (\alpha_0 - \beta_0)/\beta \\
& \quad \pi_y = (u - v)/\beta
\end{align*}

(17.3)

Since $w_x$ and $w_y$ are linear in $u, v$, they are also normally jointly distributed. Their means are

\begin{align*}
E w_x &= (E u - E v)/(\beta - \alpha) = 0; \quad \text{and similarly} \\
E w_y &= 0.
\end{align*}

Therefore by (17.2) $E x = \pi_x, \quad E y = \pi_y$, hence $\pi_x, \pi_y$ can be determined from our sample of $(x, y)$: $\pi_x, \pi_y$ are simply the averages of $x, y$ respectively.

Further, since $w_x = x - \pi_x = x - E x/\pi_x = y - E y$, the covariance matrix of $w_x, w_y$ is the same as that of $x, y$:

\begin{align*}
(17.4) & \quad \begin{pmatrix}
E w_x^2 & E w_x w_y \\
E w_x w_y & E w_y^2
\end{pmatrix} = \begin{pmatrix}
\sigma_{xx} & \sigma_{xy} \\
\sigma_{xy} & \sigma_{yy}
\end{pmatrix}
\end{align*}

and is therefore known from the sample.

Exercise 17.1. Is structure (17.1) identifiable? We shall use a method analogous to one of the two methods studied in Lecture 7, viz., the method of reduced form. Can we determine the 7 structural parameters: $c, \beta, \alpha_0, \beta_0, \sigma_{uu}, \sigma_{uv}, \sigma_{yy}$ if we know the 5 reduced form parameters: $\pi_x, \pi_y, E w_x^2, E w_y^2, E w_x w_y?$

\*Since $E w_x = E x = E y = E u = 0$, there are here no reasons for using the symbol $E$ instead of $\sigma$, except typographical ones.
Two of the relations between the known and the unknown parameters are given by (17.3), left-hand column of equations. From the right-hand column of equations in (17.3) we can derive, by multiplying and taking expected values, the following three relations:

\[
\begin{align*}
\bar{E}_{w_x}^2 &= (\sigma_{uu} + \sigma_{vv} - 2\sigma_{uv})/(\beta - \alpha)^2 \\
\bar{E}_{w_y}^2 &= (\beta^2 \sigma_{uu} + \alpha^2 \sigma_{vv} - 2\alpha \beta \sigma_{uv})/(\beta - \alpha)^2 \\
\bar{E}_{w_x w_y} &= (\beta \sigma_{uu} + \alpha \sigma_{vv} - (\alpha + \beta) \sigma_{uv})/(\beta - \alpha)^2
\end{align*}
\]

We thus have 5 independent relations, 2 too few to determine the 7 unknown structural parameters. Structure (17.1) is therefore not identifiable.

**Exercise 17.2.** The structure is as in (17.1) but there exists the following additional information: disturbances of demand and of supply are independent; and those of supply have a variance \(k\) times as large as those in demand. That is,

\[(17.6) \quad \sigma_{uv} = 0 \quad (17.7) \quad \sigma_{vv} = k \sigma_{uu}.
\]

(17.6), (17.7) supply two more equations, independent of each other and of the five equations (17.3) (left-hand pair) and (17.5). This makes the 7 structural parameters determinable from the 5 parameters of the reduced form, and therefore from the observations. Specifically, (17.6), (17.7) permits to simplify (17.5) into

\[
\begin{align*}
\bar{E}_{w_x}^2 &= (1 + k) \sigma_{uu}/(\beta - \alpha)^2; \quad \bar{E}_{w_y}^2 = \sigma_{uu}(\beta^2 + k\alpha^2)/(\beta - \alpha)^2; \\
\bar{E}_{w_x w_y} &= \sigma_{uu}(\beta + k\alpha)/(\beta - \alpha)^2,
\end{align*}
\]

where \(k\) and the expectations on the left-hand sides are known, and \(\alpha, \beta, k\) are three unknowns. After having solved for these, we substitute for \(\alpha, \beta\) into (17.3) and (17.7) and solve for \(c_0, \beta_0, \sigma_{vv}^2\).

Is there economic justification for (17.6), (17.7)? (17.6) means that the supply of the commodity at a given price fluctuates for reasons independent of those that affect the demand at a given price. For example, weather affects harvest but not tastes. This may or may not be true for a given commodity. (17.7) means, with \(k > 1\), that supply at a given price fluctuates more violently than demand at a given price. We can hardly know \(k\) precisely but may have an idea of its range, and therefore can determine the limits between which the structural parameters \(\alpha, \beta\), etc. may lie, by giving \(k\) different values.

From this example we have learned how (1) to find whether a structure is identifiable; (2) if it is identifiable, to determine the structure from the observations.

**Exercise 17.3.** Replace (17.7) by

\[(17.9) \quad \alpha = 1\]
(This information may have been supplied, for example, by a study of family budgets). Is the following structure identifiable: (17.1), (17.6), (17.9), with the understanding that \( u, v \) are distributed normally with zero means?

Now (17.3) (left-hand pair) becomes two equations in three unknowns \( (\beta, \alpha_0, \beta_0) \), permitting to express \( \alpha_0 \) and \( \beta_0 \) as a function of \( \beta \) (and of \( \pi_x, \pi_y \)); and (17.5) gives three equations in three unknowns \( \beta, \sigma_{uu}, \sigma_{vv} \). Thus the structure is identifiable.

**Exercise 17.1.** The structure is given by (17.1), (17.6), (17.7), and (17.9).

This structure is over-identified: we have now one equation more than there are unknown structural parameters. The 7 parameters can be found by using any 7 of the 8 equations, since these equations must be consistent. (This will cease to be the case when we shall drop our assumption of infinite sample and shall seek the estimates of parameters.)
Lecture 18.

Exercise 18.1. Find which parameters (if any) of the following structure are identifiable, and show how to determine them from observations.

Demand: \[ y_1 + \alpha y_2 + z = u \]

Supply: \[ \beta y_1 + y_2 = v, \]

where \( u_1, u_2 \) are jointly normal with zero means, and \( z \) is exogenous.

Reduced form: \[ y_1 = \pi_1 z + w_1 \]
\[ y_2 = \pi_2 z + w_2, \text{ where} \]

\[ \pi_1 = \gamma/(\alpha \beta - 1) \]
\[ \pi_2 = \beta \gamma/(\alpha \beta - 1) \]

\[ w_1 = (\omega - u)/(\alpha \beta - 1) \]
\[ w_2 = (\beta u - v)/(\alpha \beta - 1) \]

Again both \( w_1, w_2 \) are linear in \( u, v \) by (18.2), and therefore normal; and \( \bar{E}w_1 = \bar{E}w_2 = 0. \) Further \( u, v \), and therefore also \( w_1, w_2 \) are independent of the exogenous variable \( z \). Hence \( \pi_1, \pi_2 \) have the least squares property and can be determined from the observations on \( y_1, y_2, z \) by using "normal equations." Then the variance \( \bar{E}w_1^2 = \bar{E}(y_1 - \pi_1 z)^2 \), i.e., the sum of squares of residuals can also be determined, as can be \( \bar{E}w_2^2 \) and \( \bar{E}w_1 w_2 \). We thus know \( \pi_1, \pi_2, \bar{E}w_1^2, \bar{E}w_2^2, \bar{E}w_1 w_2 \). The unknowns are \( \gamma, \beta, \sigma \), \( \sigma_{uu}, \sigma_{vv}, \sigma_{uv} \). From (18.1), (18.2) we find \( \beta = \pi_2/\pi_1 \).

We can also obtain three equations analogous to (17.5) and involving the \( \sigma, \beta \), \( \sigma_{uu}, \sigma_{vv}, \sigma_{uv} \), i.e., still four unknowns. A glance at the original supply equation shows, however, that \( \sigma_{vv} \) can be determined (once \( \beta \) is known), \( \sigma_{uv} \) expectation of the squares of "residuals" \( \bar{E}((\beta y_1 + y_2)^2 = \beta \gamma_1 y_1 + \gamma_2 + 2 \beta \gamma_1 y_2 \). 

Exercise 18.2. Discuss the identifiability of

\[ \begin{cases} y_1 + \alpha y_2 + z_1 = u \\ \beta y_1 + y_2 + z_2 = v, \end{cases} \]

where \( u, v \) are jointly normal with zero means, and \( z_1, z_2 \) are exogenous. In this case we shall apply the "method of weighted sum," analogous to that discussed in Lecture 7. Multiply the two equations by, respectively, \( m_1 \) and \( m_2 \), and add; then by, respectively, \( n_1 \) and \( n_2 \), and add:
\[ y_1(m_1 + m_2 \beta) + y_2(m_1 \alpha + m_2) + m_1 y_1 z_1 + m_2 y_2 z_2 = m_1 u + m_2 v \]
\[ y_1(n_1 + n_2 \beta) + y_2(n_1 \alpha + n_2) + n_1 y_1 z_1 + n_2 y_2 z_2 = n_1 u + n_2 v \]

The new pair of equations lacks the following characteristic of the original one: only one of the \( z \)s was present in each original equation. Therefore, variables that obey (18.5) do not have the same distribution as variables that obey (18.4). Hence, (18.4) is identifiable, being the only structure compatible with the observed distribution of the variables \( y_1, y_2, z_1, z_2 \).

The parameters \( \alpha, \beta, y_1, y_2, \sigma_{uu}, \sigma_{vv}, \sigma_{uv} \) can all be determined by first determining the reduced form

\[ y_1 = \pi_{11} z_1 + \pi_{12} z_2 + w_1 \]
\[ y_2 = \pi_{21} z_1 + \pi_{22} z_2 + w_2 \]

since the \( z \)s are non-correlated with the \( w \)s, the \( \pi \)s can be obtained by minimizing the sum-squares \( \Sigma(y - \pi_{11} z_1 - \pi_{12} z_2)^2 \) and \( \Sigma(y - \pi_{21} z_1 - \pi_{22} z_2)^2 \); these sum-squares divided by the number of observations are (in an infinite sample) equal to \( \Sigma w_1^2 \), \( \Sigma w_2^2 \); and \( \Sigma w_1 w_2 \) can also be found, as the "mean-product" of the residuals. We have thus determined \( 4 + 3 = 7 \) parameters of the reduced form, each of which can be expressed as a function of some of the unknown seven parameters of (18.4), viz., \( \alpha, \beta, y_1, y_2, \sigma_{uu}, \sigma_{vv}, \sigma_{uv} \). That these equations are sufficient not only in number but also in variety (i.e., are independent) was already shown when we found (by the weighted sum method) that the structure was identifiable.

In the examples of this and the previous lecture, we have used two kinds of "a priori information": (1) the presence and absence of certain variables in certain equations, i.e., some coefficients are zero; (2) properties of the disturbances. For case (1) (and for the more general case of any linear restrictions upon the coefficients belonging to the same equations) the sufficient and necessary conditions for identification have been worked out, based on the "weighted sums" method, i.e., on the invariance of the distribution of observations under linear transformations. The mathematics of case (2) have been developed for the special sub-case of non-correlation between all \( G \) disturbances (as in (17.6) where \( G = 2 \)).

In Exercises 17.1, 17.2 no exogenous variables were present, and identification had to be provided from the knowledge of properties of disturbances. Richer opportunities for identification arise when the system contains exogenous variables, as in examples 18.1, 18.2. These opportunities are enhanced by the fact that, under certain conditions, lagged endogenous variables can be treated in the same way as exogenous ones—the subject of the next lecture.
Lecture 19.

See Buffalo lectures 7 and 11-17 on concepts of statics and of (1) comparative statics; (2) dynamics and comparative dynamics; (3) solution of an equation, or a system of equations; (b) solution of a differential or difference equation, or of a system of such equations; (5) the "path" (or "motion") described by an economic variable through time, in the non-stochastic case.

In the static stochastic case, the path of each endogenous variable will depend not only on the values that the exogenous variables (z) take over time, and on possible changes in the structural coefficients (c) but also on the values taken by the random disturbances (u). In other words, an endogenous variable will describe a random path, whose properties depend on the parameters of the distribution of the \( u \)'s; i.e., if the \( u \)'s are normal, on the covariance matrix \( \Sigma \).

In the dynamic non-stochastic case (see Buffalo Lectures), each endogenous variable will, in general, change through time even though the \( z \)'s and \( c \)'s remain unchanged. The dynamic properties (i.e., parameters of the path function) such as the dampening ratio depend on the values of the \( c \)'s and \( z \)'s. Hence, the practical importance of "comparative dynamics": the dampening ratio of, say, the real income is a "degree of stability" or, better, of the shock-absorbing capacity of an economic system, and is an important factor of welfare. The "social utility function" \( U \) in (1.12) may be said to depend on the dampening ratio—say, \( D_y \)—of the real income \( y \); and also on other properties of the path \( y \), e.g., on the value—say \( y_\infty \)—which \( y \) tends to approach in time. (We have assumed that the difference equations constituting the system yield a damped oscillatory solution). Hence, the importance of evaluating, say, \( \partial D_y / \partial z \) or \( \partial y_\infty / \partial z \), where \( z \) is, for example, a certain tax which has undergone changes in the observed past; or of \( \partial D_y / \partial \theta \) or \( \partial y_\infty / \partial \theta \) where \( \theta \) is a certain tax which is to be introduced for the first time. As on the Diagram in Exercise 1.2, the second type of question requires the determination of the structural constants that prevailed during the observation period.

In the dynamic stochastic case, structural constants include the distribution parameters of the disturbances; for example, the variances and covariances of the successive random variables \( u(9), u(8), \ldots \) in the following system involving only one endogenous variable \( y \), occurring in each equation at two time-points (9 means "year 1949," etc.):

\[
\begin{align*}
\begin{cases}
y(9) + \sigma y(8) + 0 & = u(9) \\
0 & + y(8) + \sigma y(7) + 0 = u(8) \\
0 & + 0 & + y(7) + \sigma y(6) = u(7) \\
& & \ddots & \ddots \end{cases}
\end{align*}
\]

\( u(9) \) is a different random variable from \( u(8) \), etc. If the \( u \)'s were zero, we could say that \( y(8) \) is exogenous to the first equation, since it determines \( y(9) \) but is not determined by it, etc. (as shown by the zeros). We could say that "past economic history" is exogenous, like meteorology. But with random disturbances present, this need not be the case: these disturbances may be correlated—e.g., an
excessively thrifty behavior of one year may be carried on next year by "inertia" (positive correlation) or may cause an opposite mood (negative correlation). Suppose, however, that the disturbances \( u(9), u(8), \ldots \) are indeed distributed independently. More precisely, consider the following propositions about the disturbances:

(A) \( \varepsilon_u(9) \equiv \varepsilon_u(8) \equiv \ldots \equiv 0 \) (across years; this is a definition, not an assumption)

(B) \( \varepsilon_u(9)u(9) = \varepsilon_u(8)u(8) = \ldots = \sigma_{uu} \) (common variance)

(C) \( u(9), u(8), \ldots \) are distributed independently of each other:

\[
Pr(u(9)) = Pr(u(9) | u(8)) \tag{19.2}
\]

(If, in addition, normality is assumed, it follows that the disturbances occurring in any two time-points are non-correlated, and that each has the same distribution.)

Condition (C) repeats for the lagged endogenous variables (a single one in this particular example (19.1)), while condition (16.5) stated for exogenous variables. (16.5) was a part of the definition of exogenous variables and of disturbances of the "non-economic world," and resulted in the proposition that exogenous variables can always be treated as non-random ones, and in the least-squares property of the reduced form coefficients. (19.2) is an assumption, not a definition. It is not a necessary property of the successive disturbances in (19.1). But if these disturbances have this property, then indeed the lagged endogenous variables can be treated as if they were exogenous ones. We can then call the non-lagged endogenous variables "jointly dependent" while the lagged endogenous ones, together with exogenous (whether lagged or not) ones may be called "predetermined." Assumption (19.2) says that past economic history does not depend on the present, but does determine the present; just as weather is not determined by prices but determines them:

\[
(19.3) \quad \underbrace{y(1), \ldots, y(t)}_{\text{jointly dependent}} \text{ } \quad \underbrace{y_{1(t-1)}, \ldots, y_{G(t-1)}, y_{2(t-2)}, \ldots, y_{G(t-2)}, \ldots, x_{1}, \ldots, x_{K}}_{\text{predetermined}}
\]

This classification is due to Koopmans. As already hinted at the end of Lecture 17, this extension of properties of exogenous variables to lagged endogenous ones enriches the opportunities for identification. For example, consider a perishable farm product: demand = supply, and there is a lag between planting and selling:

\[
\begin{align*}
\text{Supply: } \quad x(t) &= \omega(t-1) + \sigma_0 + u(t) \\
\text{Demand: } \quad y(t) &= \beta(t) + \beta_0 + v(t)
\end{align*}
\]

where \( u(t), v(t) \) are normally distributed random variables, and for any \( t, \varepsilon_u(t) = 0 = \varepsilon_v(t), \) and \( \varepsilon_u(t)u(t) = \sigma_{uu}, \varepsilon_v(t)v(t) = \sigma_{vv}, \) and also for any \( t, t', \varepsilon_u(t)u(t') = 0, \varepsilon_u(t)v(t') = 0. \) These are the conditions (A), (B), (C) above, applied to the case of two equations; because of these conditions, we can drop the
t-superscript over the u and v. (But note that \( \xi u(t)v(t) = \sigma_{uv} \) (say) need not be zero.) Rewrite (19.1), ordering the variables as in (19.3):

**Demand:** \( 3u(t) - p(t) + 0 + \xi 0 = x \)

**Supply:** \( x(t) + 0 + qp(t-1) + c = u \)

No exogenous variables are present, but \( p(t-1) \) is a predetermined one; and the structure is identifiable.

If the time-lag is small, the assumption (C) becomes less plausible; hence, it is not possible to make an otherwise non-identifiable system identifiable by confining up a tiny harmless extra time-lag.

It has been said often that since (1) in an economic time series each value usually depends on the preceding ones, while (2) "the theory of probabilities is based on the model of independent successive drawings from an urn," therefore (1) "probability theory cannot be applied to economics."

Pleasure (1) is correct, but (2) is both irrelevant and incorrect. It is irrelevant in the sense that even if (1) were true, still (1) and (2) would not imply (3). In fact, we have assumed in (19.2) that successive disturbances are independent but the time series of the observable economic variable \( y(0), y(1), \ldots, \); in our case) does not consist of independently distributed variables: \( y(0) \) clearly depends on \( y(1) \) and therefore also on \( y(0), etc. \), by (19.1).

Nor is (2) correct. The theory of probabilities has been generalized a long time ago to, say, drawings of cards without replacement. More complicated cases are naturally more difficult. Without assumption (19.2), mathematical problems arise that are unsolved but not, in principle, insoluble.

With assumption (19.2), these difficulties vanish (and identification is more often present than it would be otherwise); but the "auto-correlation" of each economic time series is not hereby disregarded.
So far, an infinite sample was assumed. Therefore the distribution of the observations, \( \text{Pr}(y|z) \) (where \( z \) now stands for all the predetermined variables) could be determined exactly. For example, each of the population means \( \mathbb{E}y_g | z \)
\((g = 1, \ldots, G)\) is approached by the corresponding sample means \( (y_g^{(1)} + y_g^{(2)} + \ldots + y_g^{(T)})/T \) as \( T \), the number of observations, "approaches infinity" (i.e., as \( T \) grows beyond any bounds). Furthermore, if the regression equation \( \mathbb{E}y_g | z = R(z) \) is linear, \( \mathbb{E}y_g | z = \pi_0 + \pi_1 z_1 + \ldots \), then the regression coefficients \((\pi)\) have the least-squares property and can therefore be determined exactly from the observations; the same is true of the variance of the residuals \( v_g = y_g - R(z) \). In general, the parameters of the reduced form can be determined exactly. On the other hand, even with an infinite number of observations, the structural parameters \((a, \Sigma)\) can be determined only if they are identifiable.

As the actual sample is never infinite, the distribution parameters of the \( y \)'s cannot be exactly determined. They can be only estimated. Hence, the parameters of the reduced form can only be estimated. The structural parameters can be estimated, provided they are identifiable. Each estimate is a function of observations and therefore itself a random variable. We write

\[
a, b, c, \beta, p, s_{uu}, S, \text{ for the estimates of } a, \beta, \gamma, \nu, \sigma_{uu}, \Sigma.
\]

Each of the estimates has its distribution: if the sample is repeated, \( b \) takes different values with certain probabilities. If \( b \) is known to be normally distributed, its distribution is characterized by \( \mathbb{E}b \) and \( \sigma_{bb} \).

Instead of obtaining \( b \), an estimate of \( \beta \), one often estimates "confidence limits"—say \( b' \) and \( b'' \)—such that in repeated samples \( b \) will fall outside of these limits with a certain probability, say 5%. The confidence limits are themselves random variables. If \( b \) is normally distributed, one can obtain its confidence limits from the estimates of \( \mathbb{E}b \) and of \( \sigma_{bb} \) (using the so-called "t-distribution").

For jointly distributed estimates—say the estimates of all structural parameters, or the estimates of future values of all \( y \)'s— one obtains from the observations the distribution parameters such as means, covariances, etc.; or one obtains a multi-dimensional "confidence region."

One can classify all estimates of a given parameter by asking either

1. how the estimate is obtained: "least-squares estimates," "maximum likelihood estimates," etc.

2. whether the estimate is "useful" ("good") in some particular respect: "unbiased," "accurate," etc.

Ultimately the "usefulness" of a given estimate formula is measured by the damage which its application is likely to inflict upon the relevant utility; the
profit of a firm, revenue of the Treasury, welfare of the community. Since the estimate is a random variable, so is the damage due to its application; and so is, say, the profit of a firm that applies this estimate in making its decisions. But it can decide to maximize the expected value (= long run average) of this profit. Then the best estimation formula is the one that maximizes the expected value of the profit. (Compare Lecture 14, remarks on Problem 13.1.) In many cases, this property will be possessed by an estimate that has the following two properties: "lack of bias" and "minimum variance." Let \( \hat{\beta} \) be an estimate of \( \beta \). Then bias = \( \beta - \hat{\beta} \). The variance \( \sigma_{\hat{\beta} \hat{\beta}} \) is a measure of "inaccuracy" of the estimate.

**Markov's Theorem.** Let \( y = \eta_0 + \eta_1 z_1 + \ldots + \eta_K z_K + v \), where \( v \) is a random variable, statistically independent of the \( z \)'s, and \( E v = 0 \). We know that then the \( \eta \)'s have the least squares property (Lecture 15). Consider a sample of \( T \) observations on \( y, z_1, \ldots, z_K \) (i.e., \( T \) points in a \( K + 1 \)-dimensional scatter diagram) and find numbers \( p_0, p_1, \ldots, p_K \) such that the sum of squares of the \( T \) residuals

\[
\sum_{t=1}^{T} (y(t) - p_0 - p_1 z_1(t) - \ldots - p_K z_K(t))^2 \equiv \sum_{t=1}^{T} (v(t))^2
\]

be a minimum. Then the \( p \)'s have the following properties:

1) each \( p \) is a linear function of the observations on \( y \); for example, if \( K = 1 \), then \( p_1 = \Sigma y(t) z_1(t) / \Sigma (z_1(t))^2 \) (the well-known "estimate of the regression coefficient of \( y \) on \( z_1 \)) = \( y(1) \cdot z_1(1) / \Sigma (z_1(t))^2 + y(2) \cdot z_1(2) / \Sigma (z_1(t))^2 + \ldots \)

2) each \( p \) is an unbiased estimate of the corresponding \( \eta \);

3) each \( p \) is the "most accurate" (or "best") linear unbiased estimate of the corresponding \( \eta \), in the sense that it has a minimum variance.

This justifies estimating the reduced form coefficients (\( \eta \)) by the least-squares method, regardless whether the residual \( v \) is distributed normally or not. If it is distributed normally, then the least squares estimates of \( \eta \), viz., \( p \), are also "maximum likelihood estimates" of \( \eta \), in the following sense. Write the probability densities of the \( T \) sample values of the residuals \( v(1), \ldots, v(T) \):

\[
\sigma^2 = \text{variance of } v \text{ and the constant } c = 1/\sigma^2 = \frac{1}{\sigma^2}. \text{ Since } v(1), \ldots, v(T) \text{ are independent, the probability density of their joint occurrence equals the product of the } T \text{ expressions in (20.2). This joint probability density is called the "likelihood" of the joint occurrence of the random observations in question, (in our case, of the values } v(1), \ldots, v(T) \text{). This likelihood is thus again an exponential, with an exponent}
\]

\[
e^{-\frac{(v(1))^2}{2\sigma^2}}, \ldots, c = e^{-\frac{(v(T))^2}{2\sigma^2}},
\]
\[ -\frac{T}{2} \sum (v(t))^2/2\sigma^2; \]

thus the likelihood is at its maximum when (20.1) is at its minimum. The least squares estimates \( \hat{\beta} \) of \( \beta \) are thus maximum likelihood estimates, if \( v \) is normally distributed. In this case the maximum likelihood estimate is therefore the most accurate unbiased linear estimate.

In general, maximum likelihood estimates do not have these strong useful properties. Instead, they often have weaker useful properties; they often are unbiased and accurate in an "asymptotic sense" (or, crudely, "in large samples"). They are "consistent" in the sense that the difference \( |b - \beta| \) tends to vanish with increasing probability as \( T \) increases; they are "efficient" in the sense that the variance \( \sigma_{\hat{b}}^2 \) tends to become a minimum (i.e., smaller than the variance of any other estimate) as \( T \) increases.

A structural equation including more than one endogenous variable does not satisfy the requirements of the Markoff theorem because each of the \( y_i \)'s is a function of the random disturbances \( (u_i) \). If we have, say,

\[ y_1 = \gamma y_2 + \gamma z_1 + u_1 \]
\[ y_2 = \beta y_1 + \gamma z_2 + u_2 \]

then, in the first equation, \( u_1 \) is statistically independent of \( z_1 \) but not of \( y_2 \). Also, the maximum likelihood estimate of, say, \( \gamma \), is not identical with the least squares estimate of \( \gamma \) obtained by regarding the first equation as a regression equation (which it is not).

On the other hand, it is possible to obtain the maximum likelihood estimates of \( \alpha, \beta, \gamma_1, \gamma_2, \sigma_{u_1}^{2}, \sigma_{u_2}^{2} \), and of the distribution parameters of these estimates. In most cases, these estimates are "consistent" and "efficient."

One way to obtain the maximum likelihood estimate (say, \( \alpha, \beta, \gamma_1, \gamma_2, \sigma_{u_1}^{2}, \sigma_{u_2}^{2} \)) of the structural parameters (say, \( \alpha, \beta, \gamma_1, \gamma_2, \sigma_{u_1}^{2}, \sigma_{u_2}^{2} \)) is from the estimates of the reduced form coefficients \( \Gamma \) and of the covariance matrix (say, \( \Lambda \)) of the corresponding residuals (say, \( p, W \)). In the previous lectures, relationships between \( \alpha, \beta, \gamma_1, \gamma_2, \sigma_{u_1}^{2}, \sigma_{u_2}^{2} \) and the information still exist. It can be shown that the same relationships exist between corresponding maximum likelihood estimates (\( \alpha, \beta, \gamma_1, \gamma_2, \sigma_{u_1}^{2}, \sigma_{u_2}^{2} \)) and (\( \pi, \lambda \)). If there are exactly as many relationships as are unknown structural parameters, we get a unique solution. If, on the other hand, the structure is over-identified, we can discard any of the supernumerary relationships (i.e., disregard some of the exogenous variables) and obtain a maximum likelihood estimate with respect to the information still used; but depending on which information we have discarded, we can get a more or a less efficient (accurate) estimate. Again, the maximum likelihood method has been applied to guide us in selecting which particular relations have to be used to insure highest efficiency.

An important practical result is the possibility of estimating a single equation of the model (say, demand equation) without knowing the nature of other equations, provided one knows some (not all) variables that enter those other equations.
Unfortunately, economic annual time series are usually too short ($T = 20$ or $25$, say) to make these advantages really significant. It is important, therefore, (1) to develop and learn to use cross-section data; (2) to learn to use quarterly and monthly series, thus giving up the convenience of treating lagged endogenous variables as predetermined (Lecture 15); (3) to develop the statistical theory of short time series, i.e., to find estimates which may not have "asymptotic" properties but can most usefully serve to predict the future values of the $y$'s (or of some politically important function of the $y$'s) on the basis of a small number of successive observations.

Note on observation errors. Throughout the course, the stochastic nature of economic observables was ascribed entirely to "random disturbances" in the equations representing the joint effect of a large number of non-measurable causes upon each relationship between variables. Another phenomenon is that of "errors of observation" in each variable also representing the effect of a large number of non-measurable causes upon the measurement of each variable. Some econometricians have, conversely, paid most attention to observation errors, disregarding the "disturbances" in the equations. We had no time to consider this side of the problem. Naturally, if both the errors and the disturbances are considered, the theory becomes more complicated. For example, even identification becomes less likely, because additional unknown parameters have now to be determined, viz., the covariance matrix of the measurement errors, (in addition to the covariance matrix $\Sigma$ of the disturbances).
Econometrics 3104/3140
Spring 1949

Problem 1. Formulate an equation to solve all 4 problems first, omitting the questions (III) in problem 2 and 4. Answer the questions (III) if time remains.

Problem 1. The quantity $x$ and the price $p$ of a perishable farm product (each measured at the population mean) are determined as in the following model (subscripts $t$ for time):

1.1. Demand: \( x_t = \alpha p_t + u_t \)
1.2. Supply: \( x_t = \beta y_{t-1} + v_t \)
1.3. The disturbance $u_t$ is not autocorrelated; nor is $v_t$.

Show (I) How to estimate $\alpha$ for a long time series?
(II) What other structural parameters are present?
(III) (If time remaining): How would you estimate those other parameters?

Problem 2. The model of the previous problem is modified as follows:

2.1. Demand: \( x_t = \alpha p_t + u_t \) (not autocorrelated).
2.2. Price fixation: \( p_t = p_t^* \); a level fixed every year by decree.

Show (I) How to estimate $\alpha$ and $\sigma_{uu}$?
(II) Is the estimate of $\alpha$ the same as in the previous problem?

Problem 3. National income $y$, consumption $c$, and annual (saving) investment $i$ are all measured in dollars of constant purchasing power and:

3.1. $y = \alpha y + \beta i + u$;
3.2. $\Sigma u = 0$;
3.3. $i = y - c$ (an identity);
3.4. $i$ is exogenous.

Show (I) How to estimate $\alpha$, $\beta$, and $\sigma_{uu}$ from a long time series of data on $y$, $c$, $i$.
(II) Suppose $y$, $c$, $i$ denote the income, consumption, and saving of an individual family which can control its savings but not its income. How does this modification affect the model and the estimation procedure from a time series of family data, or from a survey of a large number of families?

Problem 4. A survey of very large number $T$ of firms belonging to the same industry but located in places with different wage-rates $w_1, \ldots, w_T$ has been made. The price $p$ of the product is the same for all firms. Wage-rates and price are fixed independently of the firms' action. The output $Y_t$ of each firm depends on labor used only, $N_t$, according to the formula
Problem 1 (continued)

\((h.1)\) \(X_t = B_t A_t^A, \quad t = 1, \ldots, T\) (the elasticity \(A\) being the same for all firms.) Hence,

\[(h.2)\] \(R_t = \delta_t + A_t \cdot C_t, \quad t = 1, \ldots, T\) where the small letters (except for \(t\)) stand for the logarithms. Further assume that each firm pushes its output to the point where, apart from a random deviation, the ratio \((w_t/p) = R_t\) equals the labor's marginal product,

\[(h.3)\] \(R_t = (\frac{dX_t}{dt}) \cdot C_t, \quad t = 1, \ldots, T\) where \(C_t\) is a random percentage deviation. Hence

\[(h.4)\] \(R_t = A_t A_t^{-1} = B_t C_t, \quad t = 1, \ldots, T\)

\[(h.5)\] \(r_t = a + b_t + c_t + (A-1)n_t, \quad t = 1, \ldots, T\), where again: small letters indicate logarithms.

Questions: (I) How to estimate \(A\)?

(II) What other structural parameters are present?

(III) (If time remains): How to estimate those?


3. Haavelmo, T., "The Probability Approach to Econometrics" (Supplement to Econometrica, 1944).


