SALES OF NEWSPRINT IN FINLAND, 1949—1959

MODELS FOR SHORT TERM FORECASTING

PÄIVIÖ RIIHINEN

SELOSTUS:

SANOMALEHTIPAPERIN MYYNTI SUOMESSA VUOSINA 1949—1959

TUTKIMUS LYHYTJÄNTEISTEN ENNUSTEIDEN MALLEISTA

HELSINKI 1962

Suomalaisen Kirjallisuuden Kirjapaino Oy Helsinki 1962

Preface

I wish to express my gratitude for the assistance and constructive criticism that I received from so many people in preparing this study. Most of the material was provided by *Paperikonttori*, with the kind co-operation of Mr. Tauno Salmi, the sales manager. Prof. Eino Saari, my principal at the *Department of Social Economics of Forestry, University of Helsinki*, has my sincere thanks for restricting my study to its present scope and for commenting on the manuscript. I am also grateful for the comments and advice given by Prof. Viljo Holopainen and Prof. Pentti Pöyhönen in the course of the study and after their perusal of the manuscript. Mr. Jussi Linnamo has generously spent much of his time on thought-provoking discussions and perusal of the manuscript. The contribution of Dr. Seppo Ervasti, Dr. Lauri Heikinheimo, Mr. Heimo Järvinen, Mr. Lauri Korpelainen, Dr. O. E. Niitamo, Mr. Yrjö Puttonen and Mr. Olavi Riihinen also merits my deepest gratitude. I am indebted to Mr. Matti Huopaniemi for answering questions calling for an expert knowledge of journalism.

Mrs. Jenny Leander, Dr. Rein Riitsalu and Mrs. Irja Thusberg provided valuable assistance in the computations, typing and proof-reading. The figures and graphs were drawn by Mr. P. Kinnunen, and the English language checked and proof-read by Mr. R. Milton.

For the grants I have received I wish to express my sincere thanks to the *Finnish Natural Resources Foundation, Niilo Helander Foundation, Yrjö Jahnsson Foundation* and *University of Helsinki*. The *Society of Forestry in Finland* assisted the completion of this study by including it in its series of publications.

Helsinki, February, 1962.

Päiviö Riihinen

Contents

Page

Appendix Tables

Page

1 Introduction

11 Forecasting as a Scientific Undertaking

By forecasting is meant the deduction of unknown events, mainly in the future (*cf.* Theil 1958, p. 1). One of the principal aims of science is to seek invariables which can be used in forecasting. The systematic presentation of invariables — the formulation of a theory — calls for a simplification of reality: a distinction must be made between exceptional factors and essential ones in describing reality.

The invariables in a forecast are usually crystallized in a model expressed verbally, in graphs, or by mathematical symbols. The better the model complies with certain *a priori* requirements (accuracy of description or explanation, simplicity, universality), the better it is said to be. In many exact sciences the margin of error in making measurements is fairly narrow; the distribution of the error approaches some mathematically simple distribution function as the number of experiments increases. Hence forecasting in such sciences is simpler and more reliable than in those dealing with human behaviour.

It is sometimes claimed that in the social sciences the human element leaves such a wide margin of error in any conclusions drawn that forecasts are practically useless. Those who maintain this opinion, obviously consider forecasting purely as an aim in itself, and pay no attention to its role in decision making. All decisions are, in fact, made in the light of expectations, whether consciously or not. Forecasts are simply expectations expressed in an explicit form, preferably based on a model that enables a distinction to be made between systematic and random variations. To achieve this distinction, the explanatory or forecast model is given a stochastic form, which implies recognition of a randomly distributed disturbance factor. This factor provides for the »human element» mentioned above. In addition it sums up the influence of less essential factors not explicitly included in the model, and allows room for the imperfections of the theory, the aggregate nature of the variables and/or the errors of measurement (*cf.* Valavanis 1959, pp. 5—6; Niitamo and Pulliainen 1960, pp. 401—408). A stochastic explanatory or forecast model can thus be considered more practical than an exact model.

If a forecast is used only to decide between given alternatives, it is quite likely that the decision made will be different from that suggested by the forecast model. From the viewpoint of the layman the result of the forecast is then said to be »poor». If, however, it enables the decision maker to prevent an occurrence undesirable from his own viewpoint, it will still have been beneficial — at least for the decision maker.

12 The Purpose of Market Forecasts

A distinction can be made between long run and short run forecasts. In economics, »long run» refers to a period long enough to change the quantity of plant and equipment. »Short run» means a period long enough to make changes in output, but too short to alter the capacity itself (see Boulding 1955, p. 570). The actual time covered by a long run or short run analysis is therefore of minor importance in drawing this distinction; technical supply conditions, not time, are the deciding factor.

Both long and short run forecasts aim at bringing the available alternatives to the attention of the decision maker, the aim being in most cases to enable him to maximize his multidimensional welfare function by the optimum allocation of his resources.[1] But they tackle the problem in different ways: long run forecasts are used for planning conditional changes in the structure of production factors, and thus in the quantity of production (programming), or for analysing the effect of planned changes on the market (fictitious forecasting) (*cf.* Niitamo and Pulliainen 1960, p. 410). The purpose of short run forecasts is to adapt production to short term fluctuations on the market (budgeting, taking into account income transfers and restrictive conditions imposed by borrowed capital).

In forestry and the wood industry long run market forecasts have been made to analyse the following:

- Possible opening of new market outlets within a certain area (*cf.* Riihinen 1957, etc.);
- Influence of a selected forest policy on the trade balance of wood industry products (*cf. European Timber Trends* . . . 1953);
- Determination of production goals in forestry (*cf. Timber Resources* . . . 1958, Vaux and Zivnuska 1952, Riihinen 1958);
- Determination of the desirable extent and profitability of certain forest improvements (Vaux 1954);
- Expansion of the wood industries (Ervasti 1959);
- Determination of equilibrium (in terms of prices and the income level of the consumers) for forestry and the timber trade with a view to formulating forest policy (*cf. European Timber Trends* . . . 1953).

[1] This problem has been studied extensively by Theil (1958, pp. 431—529), though he makes no distinction between long run and short run forecasts.

The models for long run forecasts are conditioned by economic requirements, depending on their purpose. Here, however, the theory of long run forecasts will not be examined further.

Short run forecasts are needed for adapting the utilization of capacity to cyclical fluctuations on the market. Expenditure and revenue for a future period of production are budgeted in the light of certain expectations and restrictive conditions. Expenditure is budgeted according to the unit prices of the alternative techniques available. A forecast of revenue is largely based on a forecast of the quantity of goods that will be sold. This is often determined by »rule of thumb». Several experts, however, advocate a certain orderliness in the drawing up of a sales budget. Bartizal (1940, p. 23) recommends the following three steps:

»(1) Analysis of past results;
(2) Market analysis;
(3) Analysis of the status of each product with respect to the market and as to the profits.»

Closer inspection of points (1) — (3), however, does not reveal how the analyses should be carried out. It merely points out certain aspects that call for analysis.

Anderson *et al.* (1942, pp. 469—470) emphasize the importance of »experience» in making a sales budget. Bethel *et al.* (1956, p. 606) regard previous sales and market analysis as the most essential factors in making sales forecasts.

Quite apart from guiding management, market forecasts are of considerable assistance for the labour authorities in formulating employment policy. In the industries that supply raw materials and others with a low degree of mechanization, the labour employed is very dependent on production.

Models may exhibit alternative economic theories depending on the purpose of the market forecasts based upon them. According to the economic theory illustrated by the models, market forecasts can be classified as follows:

1. forecasts of demand
2. forecasts of supply
3. forecasts of consumption
   a. forecasts of actual consumption
   b. forecasts of potential consumption.

Forecasting demand means estimating the demand curve for a future period (*cf.* Duerr and Vaux 1953, pp. 446—447). Alternatively, two or more points along a demand curve can be determined instead of the complete curve to form a forecast of demand. In fact, the latter procedure may be the more usual (*cf. European Timber Trends* . . . 1953, *Timber Resources* . . . 1958). Apart from serving as one of the components in a forecast of consumption, forecasts of demand also provide an indication of possible prices at different levels of consumption.

Forecasting of supply involves drawing a future supply curve or determining more than one point along it.

Once both the future demand and supply curves have been estimated, the abscissa[1] of the intersection of these gives a forecast of actual consumption (*cf.* DUERR and VAUX 1953, pp. 422—423; VAUX 1954). Thus, there is no difference between the method of forecasting actual consumption and those of forecasting demand and supply. Forecasting potential consumption calls for a different procedure. Here, the consumption expected is determined on the basis of a statistical relationship between consumption and certain parameters of the economy. The demand and supply curves are not taken into account (*cf. Potential Requirements* . . . 1946). In other words, the point determined is that which corresponds to the intersection of the theoretical demand and supply curves under given circumstances. If future consumption is estimated as the abscissa of the intersection of the demand curve and a supply curve which is assumed to be a horizontal line at a given price level (*cf.* VAUX 1950), this, too, is termed a forecast of potential consumption.

Although the above classification is that used mainly for long run forecasts, it also applies in principle to short run forecasts. For practical reasons, the period of forecasting may influence the selection of the theory behind the forecast. In forecasts covering a short period of time, for instance, it is usual to keep more factors constant than in those covering a long period. Regardless of the length of the forecast period, however, the above classification holds good in interpreting forecasts made in terms of economics, irrespective of the method employed.

Since certain forecasts have failed to bear in mind the underlying economic theory, the nomenclature used in a number of studies is misleading in the light of the above classification. Thus, a forecast may be termed one of demand, even though its content and methodology seem to indicate a forecast of potential consumption (*cf. World Pulp and Paper* . . . 1954, *Possibilities for the Development* . . . 1954, *Pulp and Paper* . . . 1955, *Paper for Printing Today* . . . 1952, *Paper for Printing and* . . . 1954, *The Outlook* . . . 1957, *World Consultation* . . . 1959).

2 Purpose and Scope of the Study

Very few short run market forecasts have been made for wood industry products. Certain forecasts cover the near future (*e.g.* SHAMES 1946, *European Timber Trends* . . . 1953, ERVASTI 1959), but since they assume changing supply conditions, they must be classified as long run forecasts. One of the few short run models is GREGORY's recursive model for the hardwood flooring market (1960). Its parameters were estimated from data containing seasonal variations.

[1] Throughout this study the horizontal axis represents quantity, and the vertical axis price.

The prime object of GREGORY's model was not to forecast variations in the hardwood flooring market so much as to supply a causal explanation for them. HOLOPAINEN's (1960) object was to produce a supply equation for sawnwood exports from Finland to the United Kingdom.

Since short run forecasting is desirable both for adjusting production and for regulating the national employment policy, we shall attempt here to make such a forecast for one product and analyse the methods used in doing so.

Separate market forecasts are needed for each forest and wood industry product in order to throw light on the optimum annual production and the amount and composition of labour required for it. The purpose of this study is to develop a short run model describing newsprint sales in Finland that will combine mathematical simplicity, accuracy of description and universality. If several models are equal in all these respects, preference will be given to that which best explains both the data containing the trend and those from which the trend has been eliminated.[1] Moreover, the model will be composed preferably of variables with observed values either available from ready-compiled statistics or computable from these.

P l a c e: the study will be restricted to variations of the amount of newsprint produced in and sold to Finland.

T i m e: the period chosen will be one in which there were no official controls on newsprint consumption of a nature to restrict variations in newsprint sales.

3 Method of Study

A forecast model that enables a distinction to be made between the systematic and the random components of the phenomenon being studied can only be constructed by econometric analysis. By this is meant the application of mathematical-economic theory and statistical analysis to economic data with a view either to obtaining numerical results or testing the economic theory (*cf.* TINTNER 1952, p. 3).

A brief account will first be given of the types of econometric models, the elements in them, and the methods of estimating their parameters. This will help us select a method of estimating the parameters of the model to be used for explaining (or describing) newsprint sales in Finland or at least give us a preview of possible inconsistencies between the specification of the model and the method of estimation chosen.

The newsprint market will be described to the extent that is necessary for selecting the hypotheses required in the study, and placing them in a broader economic-theoretical context (*cf.* DUERR and VAUX 1953, pp. 18—20). We shall

[1] Trend elimination here means transforming the observations to reduce their interdependence and the collinearity, if any, between the different series.

then proceed to a detailed analysis of the factors determining newsprint consumption. First of all, a simple inventory of these factors will be given and the assumed direction of their influences considered. Secondly, the most essential factors will be selected and their time lags with respect to each other specified. We shall then enquire whether the information can be limited further (*cf. limited information method* p. 16, below) while still adhering to a logically defensible economic theory.

After these stages, we shall discuss the type of function used to describe the statistical relationship between the variable to be explained and each explanatory variable. A correlation matrix of all the variables selected will be set out. With this matrix as a guide, alternative simplified models will be specified and a series of experiments will be conducted to determine which combination of the variables best fulfills the *a priori* requirements for the model. In doing so, we shall endeavour not to lose sight of the logic of the theory; a model may seem to comply with a certain explanatory requirement because all the variables have a common linear trend or some other »explanatory» factor not explicitly considered. To avoid falling into this trap, analyses performed by means of the series including the trend will be compared with those carried out with the first logarithmic differences of the observations.

4 Types and Elements of Econometric Models

A certain general knowledge of the types and elements of econometric models is necessary to permit an understanding of the methods of estimating the parameters of those used later in this study. These have been presented in several textbooks and manuals of econometrics, from which the following summary has been adapted (as, too, has the introduction to methods of estimating parameters in Chapter 5).

An econometric model can comprise one or more equations. A single-equation model consists of one regression equation, *e.g.*

$$(1)\ x_1 = \alpha_1 + \alpha_2 x_2 + \ldots + \alpha_n x_n + u,$$

in which x_1 is the variable to be explained, $x_2 \ldots x_n$ the explanatory variables, $\alpha_1 \ldots \alpha_n$ parameters and u the random variable (disturbance factor). The purpose of such a model is to describe one aspect of the phenomenon (*e.g.*, demand or supply) even if it fails to produce identification (*cf.* p. 15, below).

Some of the variables $x_1 \ldots x_n$ are endogenous; these influence each other either simultaneously or with lags. Others are exogenous; they influence endogenous variables but are not influenced by them. Lagged endogenous and exogenous variables are together called »predetermined». — An endogenous variable may also be defined as one that is correlated with the unexplained residual (FOOTE 1958, p. 60).

A multi-equation model can either be recursive or simultaneous. The properties of recursive models have been studied mainly by WOLD (1953 and 1956) in order to create models whose parameters could be estimated by the method of least squares. Such a model might consist of the following system of equations, in which $a_{10}, a_{20} \ldots a_{mm-1}$ are parameters, $X_{1t}, X_{2t} \ldots X_{mt}$ endogenous variables, $L_{1t}, L_{2t} \ldots L_{mt}$ linear functions of the predetermined variables, and $u_{1t}, u_{2t} \ldots u_{mt}$ normally distributed random variables.

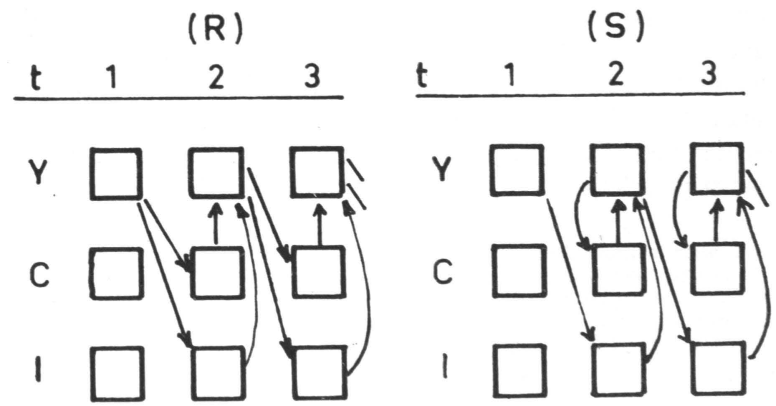
$$
\begin{aligned}
X_{1t} &= a_{10} && + L_{1t} + u_{1t}' \\
X_{2t} &= a_{20} + a_{21} X_{1t} && + L_{2t} + u_{2t}' \\
X_{3t} &= a_{30} + a_{31} X_{1t} + a_{32} X_{2t} && + L_{3t} + u_{3t}' \\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
X_{mt} &= a_{mo} + a_{m1} X_{1t} + a_{m2} X_{2t} + \ldots + a_{mm-1} X_{m-1} && + L_{mt} + u_{mt}.
\end{aligned}
$$

Once the values of the predetermined variables have been ascertained, the values of their linear functions $L_{1t}, L_{2t} \ldots L_{mt}$ are also known. If we omit the random variables $u_{1t}, u_{2t} \ldots u_{mt}$, the entire system becomes deterministic and forms a recursive causal chain. The value of X_{1t} (in the period t) can be calculated from the first equation and substituted in the second equation, which gives the value of X_{2t}. Proceeding in this way, the values of the endogenous variables in the last equation can also be calculated.

Simultaneous models also consist of systems of equations. They differ from recursive models mainly in that they contain simultaneous influences between pairs of variables and that a single equation can include more than one endogenous variable without restriction.

Models (R) and (S) illustrate the differences between recursive and simultaneous models:

$$
(R)\quad
\begin{aligned}
Y_t &= C_t + I_t \\
C_t &= f(Y_{t-1}) \\
I_t &= F(Y_{t-1})
\end{aligned}
\qquad\qquad
(S)\quad
\begin{aligned}
Y_t &= C_t + I_t \\
C_t &= f(Y_t) \\
I_t &= F(Y_{t-1})
\end{aligned}
$$



The differences between cause-and-effect relationships in these models can also be seen from the corresponding arrow schemes, which have been presented in several publications (*cf.* WOLD 1953, p. 13; HAIKALA 1956, p. 44; NIITAMO and PULLIAINEN 1960, p. 407; etc.).

5 Estimating the Parameters of a Model

The parameters of an econometric model are estimated from empirical data on the assumption that the time series representing the variable to be explained in each equation is a function of the variables explaining its systematic variation and of a random variable, as indicated by equation (1) above. The properties of the parameter estimates vary with the specification [1] of the model, the method of estimating its parameters and the properties of the residual term obtained in applying each method of estimation. The specification of the model, the method of estimation and the residual term should be such as to make the estimate

1. a maximum likelihood estimate [2]
2. unbiased (the expectation value of the estimate is equal to the true parameter value)
3. consistent (approaching its true parameter value as the size of the sample increases)
4. efficient (the estimator possesses the smallest variance of all possible estimators)
5. sufficient (including all the information on the parameters that can be obtained from the sample).

A consistent estimate is unbiased if all possible data have been used in computing it. In small samples it may be either biased or unbiased (*cf.* MOOD 1950, p. 150). In practice all the possible data are very seldom available. Preference is then given to specifications of models and methods of estimating that will produce consistent estimates, rather than ones that are known to lead to biased estimates, even in large samples. If there is only one endogenous variable in the equation and this is selected as the variable to be explained, its parameters can be estimated consistently and efficiently by the method of least squares [3], provided that the distribution of the residual term complies with certain *a priori* requirements (KLEIN 1953, pp. 80—85). For instance, the residual term must follow some probability distribution (if this is normal, the estimate will be one of maximum likelihood), and its variance must be independent of the values of the explanatory variables (FOOTE 1958, pp. 58—59). In practice, especially with small samples, only the randomness of the residual term can be tested. Apart from stipulations set on the residual term, another *a priori* requirement

[1] By specification is meant expressing an economic theory in mathematical terms.
[2] A maximum likelihood estimate is not necessarily one obtained by the maximum likelihood method.
[3] The purpose of the method of least squares is to minimize the sum of the squares of the deviations. As applied to equation (1) (p. 12, above), this means determining the parameters $\alpha_1 \ldots \alpha_n$ in such a way as to minimize the value of $\Sigma (x_1 - x_1^*)^2 = \Sigma u^2 = (x_1 - \alpha_1 - \alpha_2 x_2 \ldots - \alpha_n x_n)^2$. x_1^* is the calculated value of x_1.

for an efficient estimate by the method of least squares is that each explanatory variable should be composed of series of known (not random) values that are free of observation errors (*cf.* FOOTE 1958, pp. 58—59).

Provided that the above conditions are complied with, the method of least squares can also be used to estimate the parameters of a recursive model consistently and efficiently, one equation at a time. The parameters of simultaneous models can be estimated by the method of least squares provided that the models can be transformed into the »reduced» form, from which the values of all the non-lagged endogenous variables are calculated as functions of all the predetermined variables in the system (see FOOTE and FOX 1954, p. 44).

In studying methods of estimating the parameters of equations with several endogenous variables, most investigators resort to the concept of identifiability. Sometimes it is impossible to estimate parameters with the data available. The equation is then said to be under-identified. More often than not, the conditions for identification only comply with the formal requirements for identification: data are available on variables that, according to economic theory, are shifters of demand or supply, etc., though their statistical properties (multicollinearity) may impede identification, especially in small samples (the standard errors of the parameters may be excessive). Formal identification of a given degree for the parameters of a linear model is governed by three principles:

1. If the number of variables in a system of equations comprising the model minus the number of variables in the equation being considered exceeds the number of equations in the system less one, the economic-statistical relationship under consideration is over-identified.
2. If the above magnitudes are equal, the economic-statistical relationship is just identified.
3. If the inequality mentioned in 1 is reversed, the relationship is under-identified.

In practice, identification based on these rules is only possible providing the explanatory variables are free from collinearity. Otherwise the estimation of one or more parameters must be restricted by an *a priori* condition, — for example, by estimating the parameter for the income level from cross-sectional data in which the price is constant. The problem of identification has been dealt with extensively in econometric literature (*cf.* KLEIN 1953, pp. 92—100; TINTNER 1960, pp. 44—46, 238—243; VALAVANIS 1959, pp. 85—106), and requires no further examination here.

Although the transformation into reduced form of a model with several non-lagged endogenous variables complies with one of the fundamental requirements for using the method of least squares, it may be difficult to obtain a normal distribution for the residual term. When this distribution deviates from normal, one of the prime requirements for obtaining estimates of maximum likelihood for the parameters is lacking. KLEIN (1960, p. 870), too, considers that the efficiency of estimates obtained from the reduced form is debatable (*cf.* HURWICZ 1950, p. 272). Hence, use has been made of the method of maximum likelihood (though not in studies on markets for forest products). The estimates produced by this method are modes of their distributions, while those given by the method of least squares are means. Under certain conditions, estimates of maximum likelihood are consistent and efficient; they are also sufficient in cases in which there exists a function of sufficient estimation (*cf.* TINTNER 1960, p. 35).

The disadvantage of the method of maximum likelihood is the amount of sheer labour it entails. This applies particularly to the »full information method» (*cf.* KLEIN 1953, pp. 100—122). The *Cowles Commission for Research in Economics* has developed a »limited information maximum likelihood method» (CHERNOFF and RUBIN 1953, pp. 200—212; KLEIN 1953, pp. 125—133). This method presupposes the *a priori* restriction that the coefficients of the variables with unknown observed values omitted from the model are zeros. FOOTE (1958, pp. 63—64) summarizes the method as follows:

»In essence, a least squares fit is obtained between a linear combination of the endogenous variables in the equation on the predetermined variables in the equation, subject to the condition that the predetermined variables in the system but not in the equation have a minimum effect on the combination of the endogenous variables».

The estimates provided by this method are at least as consistent and efficient as those given by any other method using the same amount of information.

There are other possible methods of estimating the parameters of equations with several endogenous variables, such as the instrumental variables method (KLEIN 1953, pp. 122—125) and THEIL's (1954 and 1958) two-stage method. These and other methods do not concern us here.

6 Newsprint Sales in Finland, 1949—1959

61 Newsprint Market

611 Product and Type of Competition

»Newsprint» is taken to mean machine glazed paper weighing 52 grm./m², used in newspapers.[1] As far as the process of production is concerned, newsprint could also include supercalendered newsprint and newsprint waste. The markets for these, however, are not included in this study. Supercalendered newsprint is mainly used in periodicals and journals, and newsprint waste for wrapping, packaging, etc.

Newsprint (in our meaning of the term) is a homogeneous product, which has practically no substitutes. On the other hand, in both advertising and circulating news, other, more modern media of mass communication are valid substitutes for a newspaper.

[1] A »newspaper» is defined here as a periodical publication printed on newsprint, regardless of how often it appears. Thus there are weekly and even monthly »newspapers», though the majority are, of course, daily.

A description of competition might begin with the idea of »perfect competition», a theoretical concept developed mainly to facilitate comprehension of price formation, but hardly ever found in practice. This concept has been presented in many textbooks and manuals on economics, as well as studies on the theory of competition (*e.g.* STIGLER 1952, BOULDING 1955; see also MATTINEN 1955, pp. 18—21). It is not intended here to repeat the theory behind the concept, but to consider, first, how the newspaper market diverges from this model of perfection, and secondly, how it differs from the other extreme — *i.e.* full monopoly.

The sole distributor of newsprint in Finland is *Paperikonttori* (a department of the *Finnish Paper Mills' Association*). This is a joint sales organization run by the mills. Theoretically it is in a position to fix the price at a desirable level, the upper limit being presumably dictated by foreign newsprint prices. The seller has to bear in mind that the buyers (the newspapers) are well aware of the market situation and can import their newsprint if they are dissatisfied with the home price. Prices are fixed usually for a year at a time after negotiations between the seller and the organizations representing the newspapers. The seller takes it for granted that all the buyers are charged the same price. *Paperikonttori* considers this principle so self-evident that there is no reason to specify the motives behind it. One may, however, think of such factors as the desirability of avoiding discrimination between the political parties represented by the papers or any other restriction of their freedom of expression. In fact there is only one buyer, for the organizations representing the newspapers always act jointly at price negotiations. Thus the newsprint market virtually forms a bilateral monopoly. Price formation under such conditions has been investigated relatively little. Theoreticians in the field include ZEUTHEN (1930 and 1955), HICKS (1932), NICHOLLS (1941) and STIGLER (1952).

Quality and service competition are not considered here (for quality competition, see ABBOTT 1955, etc.). The basic idea behind this study is that, in negotiating with the organized buyer, the seller has only the price parameter at his disposal. The theoretical possibility of the seller pressuring the buyer by threats of changes and increased sales for export is irrelevant in practice, for the former is restricted to the unwritten conventions mentioned above.

612 Number and Demand for Newspapers, Demand for Newsprint

In 1949—59, the number of newspapers in Finland varied between 174 and 207 (SVT, *Tilastollinen vuosikirja*). In 1959, those with six or seven issues a week numbered 58 (62 in 1956). In 1949 there were 108 papers with at least three issues per week and in 1959 there were 96 (*Lehtien postimaksukomitean mietintö* 1959, p. 9). According to *Lehtien postimaksukomitea* (*Committee on Newspaper Postage Rates*), the ratio of newspapers to all printed matter is increasing despite the reduction in their number. This is an indication of the growth of large newspapers.

Information exists on newsprint consumption by newspapers of different size during the period of paper rationing. In 1944—48 *Helsingin Sanomat* alone used 22—23 per cent, and *Helsingin Sanomat*, *Hufvudstadsbladet* and *Uusi Suomi* together took 41—44 per cent of all the newsprint consumed in Finland (Koivulehto 1949, Appendices). With the continued growth of the large papers, these percentages may have increased since.

The newsprint market cannot be described adequately in isolation from the newspaper market. The amount of newsprint required by the newspapers equals the amount (in weight) of newspapers purchased by the public plus a very small quantity of job printing (circulars, pamphlets, etc.), printing waste, defective copies and unsold single copies.

Certain conclusions can be drawn on logical grounds from the demand function for newsprint. To do so, however, we must first examine the demand function for newspapers. — The newspaper is at the present regarded as an almost inevitable commodity of consumption. The fresher it is delivered to the consumers, the greater its value. In this respect, present-day papers differ very little from each other. On the other hand, each newspaper bears the »trademark» of its political views, and the traditions and other social characteristics it represents. Despite the non-uniformity of their prices, this tends to restrict the substitution of one newspaper for another. Other media of mass communication have not yet succeeded in replacing the newspaper. Radio, television, etc., are probably largely complementary and may be partial substitutes. The price of an annual newspaper subscription is so small compared to the reader's (or family's) annual income that he is not likely to give it up on account of an increase in price. Subscribers of more than one paper, however, may well consider abandoning one of them if prices rise. It may be concluded that the demand curve of the newspaper subscribers is rather inelastic, though it tends to shift with the income of the reader. This shift manifests itself as an increase both in circulation and in the use of advertisement space (*cf.* p. 23).[1]

From the demand function of the subscribers conclusions can be drawn regarding the demand function of the newspaper publishers for newsprint. Obviously the latter are aware of the subscribers' attitude towards subscription rates and can react to changes in newsprint prices accordingly. They can transfer at least part of any increase in newsprint price to their subscription rates without

[1] There are many references in the literature to the influence of income level. In the United States, this influence has been clearly established (see *Pulp, Paper and* ... 1957, pp. 56—58). It appears indirectly from certain sociological investigations carried out in Finland, in which newspaper subscription has served as a measure of social activity in socio-economic classifications (*cf.* Sariola 1954, pp. 26—28). Quist (1960, p. 76) established a highly significant correlation between economic status and reading interests. He explained growing interest in reading with rising wealth in terms of increased school attendance brought about by improved financial means.

seriously prejudicing the total revenue from subscriptions. However, a newspaper hardly ever bases its price merely on considerations of total revenue. In most cases an overriding incentive is to maintain or increase the paper's circulation. In 1950, the cost of newsprint accounted for 4—26 per cent of the total cost of the average newspaper, depending on its size and circulation (Berg 1952, p. 58).

The publisher can also adapt his economy to changes in the price of newsprint by altering his advertisement rates. It is not intended here to draw definite conclusions on the demand function for advertising space. This may be rather inelastic, though it is susceptible to shifts with changes in the income level. Neither shall we analyse how newspapers relate their subscription prices to their advertising rates. The general conclusion can be drawn that both subscription and advertising rates can be raised if the price of newsprint goes up. This will suffice in formulating econometric hypotheses later on in the study. No attempt will be made to establish a statistical demand curve for newsprint. In view of the above arguments it would seem to be fairly inelastic, though it is difficult to verify this conclusion against the data available. Without *a priori* restrictions, there can be no certainty as to which changes in the quantity of newsprint sold are due to shifts in the intersection of demand and supply along the demand curve, and which are due to shifts in the demand curve as a whole.

In our later selection of hypotheses for econometric analysis we shall investigate the rhythm in which the newsprint seller, the publishers, the subscribers and the advertisers make their decisions, — *i.e.* the time lags by which the decisions follow each other. The number of decision makers who affect the newsprint market is considerable. In addition to the seller, it includes some 200 newspapers, about 1.5 million subscribers and a large number of advertisers (*cf.* Berg 1952, p. 14).

613 The Supply of Newsprint

Newsprint sold in Finland forms only a small part of the newsprint produced here. Between 1949 and 1959 this ratio rose, with some annual fluctuations, from ca. 7 per cent to almost 10 per cent (Table I, p. 72). In negotiating the price for each coming year, the seller has a rough idea from previous experience of the quantity that can be sold in that year. The short run supply curve thus forms a horizontal straight line. From year to year this moves from one price level to another according to the cost level, newsprint export prices, etc. Accordingly, the seller is willing to sell as much newsprint as the publishers need in a particular year at the agreed price — knowing that the ratio of domestic sales to total newsprint output fluctuates annually by no more than 1.5 per cent (see Table I, p. 72, below).

614 Aims of the Econometric Analysis and the Underlying Hypotheses

Our econometric analysis will aim at developing a simple and, if possible, generalizable model explaining or describing the variations in the amount of newsprint produced in and sold to Finland between 1949 and 1959. There are not sufficient data on all the selected variables to permit a longer period of observation. Newsprint rationing only ended in January 1949 (*cf.* Koivulehto, p. 22). Thus the post-war period up to this year is excluded from the analysis. Since the data consist of annual items, there are 11 degrees of freedom available. This sets limits on both the number of equations in the model and the number variables in them. We shall try to express the model by a single regression equation.

The model should be capable of forecasting single points, but no effort will be made to work out a future demand curve, income elasticity, etc. Inaccuracy of individual parameters due to collinearity between the variables (see p. 32, below) and the method of estimating the parameters will thus be less significant than if we attempted to determine the net influence of a certain variable, with other variables being held constant. It will be endeavoured to reduce the statistical bias of the the parameter estimates by specifying the model to comply with certain *a priori* requirements mentioned above (*cf.* pp. 14—16).

In selecting the hypotheses for analysing annual changes in domestic newsprint sales by econometric means, we first tested a few that did not conform explicitly to economic theory. The explanatory power of the models based on these hypotheses, as measured by the square of the coefficient of multiple correlation, seemed to be more than 95 per cent. To test the logicality of these models, it was considered necessary to be able to draw an arrow scheme indicating the influences of the variables in consistence with a logically defensible economic theory. This brought out flaws in the logic of the models according to economic theory. Thus, at the cost of considerable wasted labour, the lesson was learnt that, logical hypotheses were called for, first of all and only after that a search for data on the appropriate variables.

The factors affecting the decisions made by newspapers in purchasing newsprint are the object of study. The scheme in Fig. 1 shows most of these factors, but not the rhythm of decision making or the proportional significance of each factor.

Some of the factors in Fig. 1 are not very significant in making decisions on newsprint purchases. Assuming that a newspaper — after duly circulating the news and guiding public opinion — takes profit maximization (or minimization of loss) as its secondary objective, newsprint purchases are then affected by the factors that determine the paper's total revenue and expenditure. Of these, the most significant are the ones with the greatest influence on total revenue and costs. Among the most important are those which vary with output (variable costs).

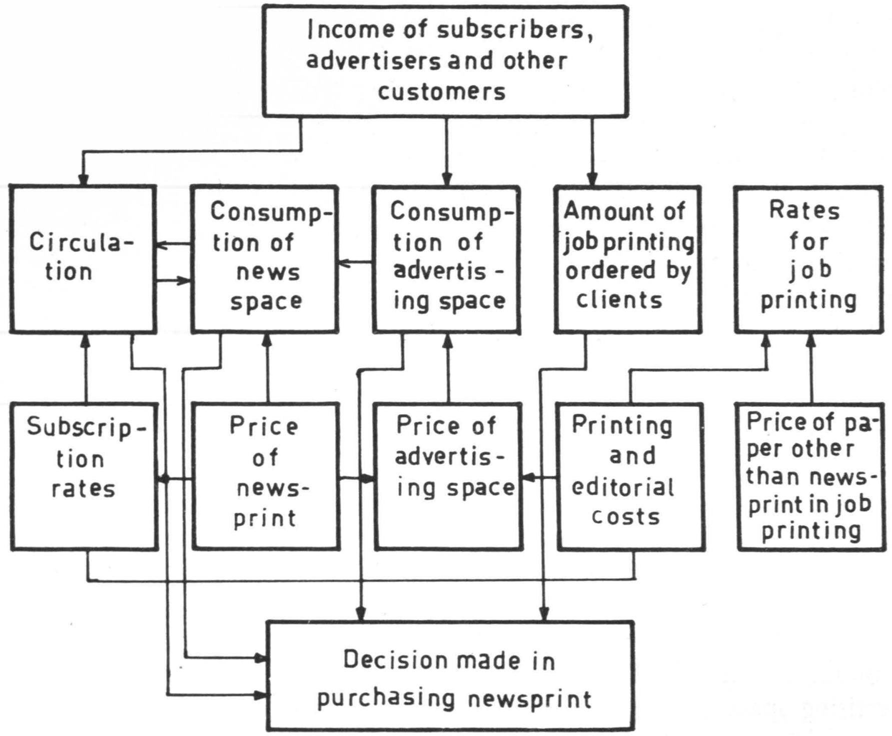


Figure 1. Hypothesis of the factors affecting decision making in purchase of newsprint. The arrows indicate the assumed direction of the influences.

A rough idea of the contribution of different proceeds and costs to total revenue and total expenditure in 1950 is given by the following percentages, which vary according to the size and circulation of the paper (Berg 1952, pp. 47—83):

Proceeds		
Newspaper sales	32—55 per cent	
Advertisements	44—67 » »	
Other revenue	.5—15 » »	
Costs		
Printing	20—30 per cent	(*f'*)
Editing	17—19 » »	(*f*)
Newsprint	4—26 » »	(*v*)
Distribution	7—17 » »	(*f'*, *v*)
Administration, maintenance, etc.	20—40 » »	(*f*)

Symbols in brackets: (*f*) = fixed; (*f'*) = fixed within certain limits, changing »in steps»; (*v*) = variable. — In the short run, only cost of newsprint is strictly variable. In 1950, the newsprint in the three biggest Helsinki papers accounted on the average for 26.3 per cent of the total cost of publishing, in 14 Finnish papers 21.8 per cent, in five provincial papers 18.8 per cent, and in three small rural papers 4.0 per cent (Berg 1952, p. 58). Even fixed costs vary in the course of time. For example, the cost of distribution depends on the level of wages, postage and other transportation rates. In the absence of adequate data, however, it is impossible to assess the influence of fixed costs. There can be no advance indication of their significant effect. In constructing the model this effect may be included in other variables — without losing degrees of freedom by adding separate variables.

After elucidating the above points we shall endeavour to select the most significant variables and to examine the time lags with which these affect each other according to the rhythm of decision making at all stages. The price of newsprint can be taken as the point of departure in this examination.

Once a newspaper publisher knows the price of newsprint, he can immediately make a decision regarding the rates for advertising space. On the other hand, the subscription rates for the year to which the newsprint price applies cannot be modified, since they will have already been decided earlier. It will only be possible to take the modified newsprint price into account in the subscription rates for the following year. Most newspaper subscriptions are annual.[1] The decisions of the advertisers are mainly influenced by the purchasing power of the public in the period affected by the decision. Hence the consumption of advertising space in a certain year is associated with the income level in the same year (in the short term, this does not apply to announcements independent of the business cycle). Circulation, on the other hand, can only be affected by changes in the purchasing power of the public with a lag of one year, because subscribers make their decision for the year *t* in the light of their income during the year *t—1* (and assume that their income will continue to be the same in the year *t*). A subscription for one year must be paid either at the end of the preceding year or at the beginning of the year in question. Fig. 2 has been drawn in view of these facts.

Symbols used in Fig. 2:

Y_{t-1} = income level of subscribers, advertisers and other customers in the year *t-1*
A_{t-1} = consumption of advertising space during *t-1*
π_{t-1} = average unit price of advertising space during *t-1*
P_{t-1} = price of newsprint in *t-1*
p_t = average (annual) subscription rate during *t*
C_t = circulation during the year *t*
$R_{A_{t-1}}$ = revenue from advertisements during *t-1*
R_{C_t} = revenue from subscriptions in the year *t*
$\bar{Q}$ = quantity of newsprint sold during the year *t*

[1] Mr. Matti Huopaniemi (*Helsingin Sanomat*) estimates the proportion of one-year subscriptions as roughly 80 per cent.

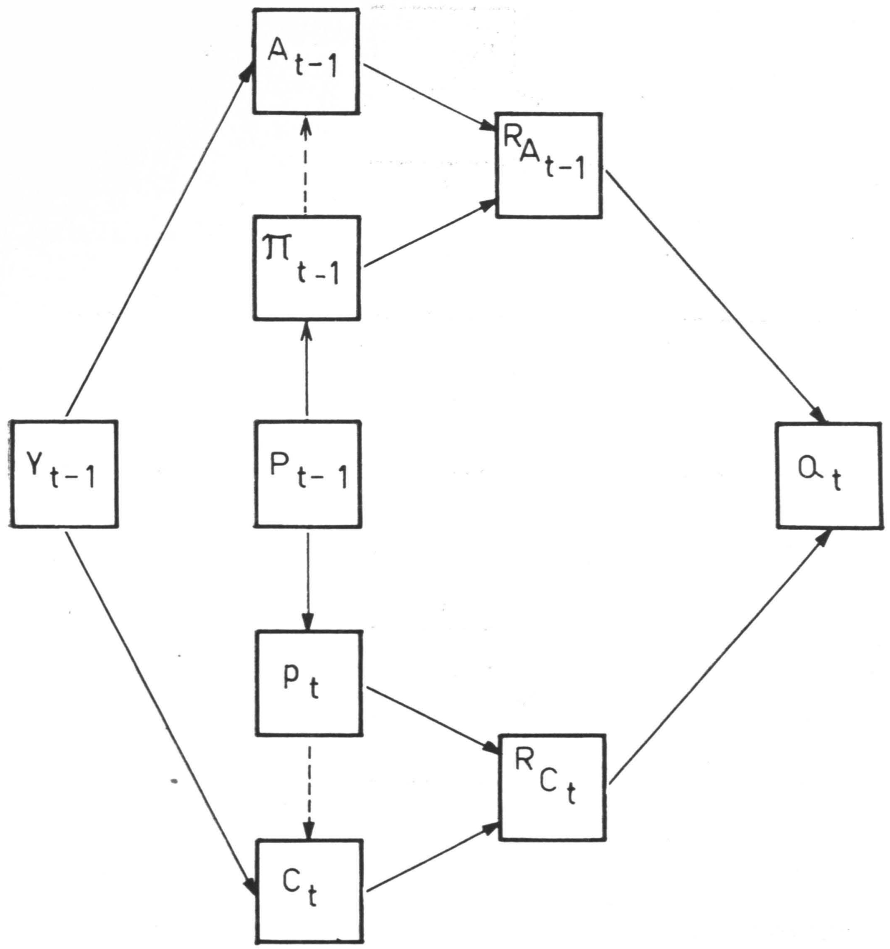


Figure 2.

The broken line arrows signify influences that cannot be verified by verbal reasoning (*cf.* p. 18, above).

If the consumption of advertising space (A_{t-1}) is inelastic in respect of price (π_{t-1}), the variable $R_{A_{t-1}}$ can be replaced by the corresponding volume series A_{t-1}. The same applies to R_{C_t} if the demand for a newspaper is inelastic. These possibilities are equally valid as hypotheses and define the object of measurement with equal accuracy. If the aim of maximizing the profit[1] of the newspapers is

[1] Profit maximization may be a less important consideration for newspapers than it is in business life in general. For instance, a newspaper may use a rise in revenue from advertisements merely to increase the amount of text and thus improve its service to the public. But even if profit maximization is not the aim, it is obvious that revenue from advertisements largely determines the size of the paper and can thus affect its circulation (*cf.* Berg 1952, p. 41). In some cases, newspapers receive outside funds.

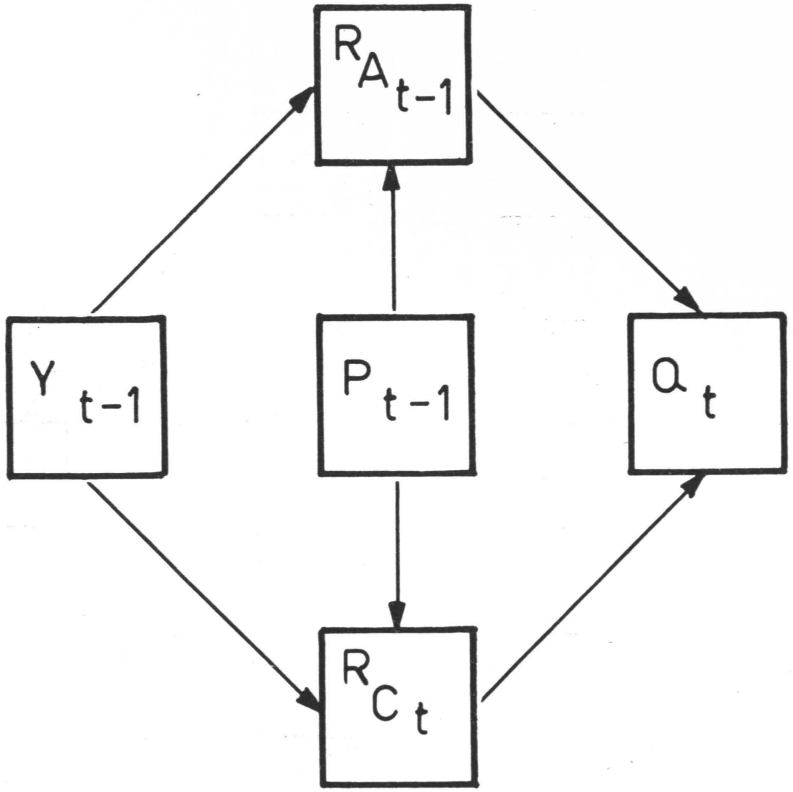


Figure 3.

taken as an alternative hypothesis (see p. 20, above), the question of inelasticity in the above prices need not be raised. In any case it seems that Fig. 2 above can be simplified into Fig. 3.

The consumption of newsprint by newspapers can also be analysed from hypotheses that are not directly related to decision making in purchasing newsprint. Thus the consumption of newsprint (Q) can be defined as a function of two factors, circulation (C) and the average size of one copy (V)[1]:

(2) $Q = f_1 (C, V)$.

Circulation, in turn, can be expressed as a function of population (H), average real income per inhabitant (Y) and average subscription rate (p):

(3) $C = f_2 (H, Y, p)$.

[1] The form of analysis implicit in equations (2) — (7) was suggested by Dr. O. E. Niitamo.

The average size of one copy is defined as function of the space occupied by news and features (N) and advertising space (A):

(4) $V = f_3 (N, A)$.

The consumption of news space can be expressed as a function of time (t) and business cycles (B):

(5) $N = f_4 (t, B)$,

and the consumption of advertising space as a function of the economic activity of the advertisers (E), that of the service industries (E'), that of their clients (E''), the average price of advertising space (π), and time (t):

(6) $A = f_5 (E, E', E'', \pi, t)$.

The consumption of newsprint per time unit (Q) can then be presented in a new form:

(7) $Q = C (N + A)$.

Equation (7) can be simplified into

(8) $Q = CZ$,

in which $Z = N + A$.

The (latent) model of newsprint consumption implicit in the above equations must, when applied, be transformed to enable each variable to be provided with a time series representing it. For example, the economic activity of advertisers and users of services (E and E'') might be described by the volume index of trade, industry and house building (T), and that of the service industries (E') by the index of unemployment (L). The variable Y could perhaps be described as the ratio of real income per inhabitant to the average subscription rate $(\frac{Y}{p})$ and time as the level of knowledge (K) (*cf.* Niitamo 1958; *cf.* also p. 18, footnote). As is the case with consumption of news space (N), all variables hardly have a series representing them. This is not even necessary, for the number of variables in an application will in any case be rather small owing to multicollinearity, among other reasons (see p. 32, below). The above exposition, however, is valid as a logical structure from which a simpler model may be obtained by combining the equations. The lags between the variables can be found or inferred from Fig. 2 (p. 23, above), assuming that newsprint consumption is described by a series showing its sales. Other aspects of application will be dealt with later on (p. 34).

62 The Most Important Experiments with Models

621 Form of Function

The form of the function used to describe the statistical relationship between the variable to be explained and each explanatory variable can be selected by logic, aided by tentative statistical experiments. In a multi-variable equation there is hardly ever only one logically defensible form of function. The form chosen is usually a compromise between a simple form that roughly illustrates the relationship and a complicated one attained after examination variable by variable. In this study, we shall utilize the logarithmic function of the form:

$$(9)\ \log x_1 = \alpha_1 + \alpha_2 \log x_2 + \ldots + \alpha_n \log x_n + u.$$

The symbols are the same as in equation (1) (p. 12, above).

A straight-line logarithmic equation can be justified here seeing that changes in variables associated with each other in economic phenomena are relative rather than absolute within a certain range. Thus, too, the relative accuracy of the model remains the same through the observation period despite a heavy rising trend. The practical advantage of a logarithmic function is that elasticities can be noted directly from the coefficients. One disadvantage is that a logarithmic function represents constant elasticities, whereas the income elasticity in newsprint consumption within a wide range can hardly be constant. If, however, the period and area covered are limited, it may be very nearly constant. This problem has been analysed by *FAO* — perhaps with data containing quite a few errors of observation — and the conclusion was reached that the dependence of paper consumption on the level of income was best approximated by the sigmoid representing the integral of the logarithmic normal distribution [1] (*World Demand* . . . 1960, pp. 97— 98). This conclusion is especially justifiable in the case of newsprint because the relative increase in newsprint consumption with rising income obviously tapers off. Paper consumption as a function of income level finally reaches saturation point. It does not fall from this level, however, because, to date, there are extremely few substitutes for paper. A second-degree parabola, would give almost the same result provided it did not reach its peak within the range under consideration. In a multi-variable equation, however, curvilinear functions, especially these mixed with straight-line functions, would necessitate a complicated and laborious procedure of computation that would hardly justify the effort by the benefits it offered.

---

[1] The equation of logarithmic normal distribution is obtained from the equation of normal distribution by introducing the independent variable in logarithmic form (see AITCHISON and BROWN 1957).

622 Material for Analysis

622.1 Variables and Series Representing Variables

The variables to be included in a regression analysis must be described by empirical series. The contents of some of these are exactly what their names imply; others include certain side influences. In econometrics it is necessary to use substitute series the meaning of which cannot be expressed briefly in words. More essential, however, than their verbal expressability is their ability to explain or describe a course of events. — To eliminate the effect of differences in the unit of measurement, all the variables will be expressed as index figures. These are set out in Table IV (p. 75).

The series describing the domestic sales of newsprint was extracted from price adjustment sheets supplied by *Paperikonttori*. Since this organization is the sole distributor of Finnish-made newsprint on the domestic market, the series is probably free from reporting errors. The variable Q_t used in the regression analysis refers to the quantity of newsprint sold in kilos per inhabitant (only those 15 years of age or more are counted). Thus the substitution of a collinear variable for population in the equation is avoided and one degree of freedom saved. The influence of population must, however, be taken into account in some form, to enable the model being constructed to be used for forecasting the total consumption of newsprint. This can be done by multiplying the forecast of consumption per 15-or-more-year-old inhabitant by the corresponding population section forecast.

A monthly index of consumption of advertising space has been published by the *Economics Department of the Ministry of Finance* in *Suhdannesarjat* since the latter half of 1948. The index figures are computed by the *Economics Department* from a sample collected by asking certain newspapers for monthly total data. The size of the sample has changed somewhat in the course of time.[1] Here, too, the consumption of advertising space (A_{t-1}) is expressed per inhabitant of 15 or more years of age. The annual data required for this study were computed in the form of arithmetic means of monthly figures. The figure for 1948 was computed from data on three Helsinki papers obtained from a publication by BERG (1952, p. 74). A check against data for a few years after 1948 showed a difference of not more than one index point between the figures from these two sources.

»Circulation» (C_t) is described as the number of copies of periodicals (including those other than daily papers) distributed by mail every year. The name given to this variable, however, cannot be used without bearing in mind its somewhat misleading content: papers distributed by mail also include journals and periodi-

---

[1] The composition of the sample has remained unchanged since 1954. It comprises total data from 19 large and 14 small newspapers.

cals, as there are no separate statistics for these. In describing circulation, the number of papers distributed by mail can only be considered correct if the ratio between the number of newspapers and that of other papers distributed by mail has remained constant the whole time. On the other hand the trend in the number of papers distributed by mail may reveal the influence of improved postal services. It was not possible, for example, to investigate whether the ratio between the number of papers distributed by mail and that distributed in towns and other densely populated communities by the papers themselves has changed significantly. Such additional study would have been an arduous undertaking, not least because the influence of population movements to towns would have had to be eliminated. In this study, the assumption is made that the publishers' own lorry transportation of papers has remained constant in its ratio to the number of copies distributed by mail. In the case of *Helsingin Sanomat*, this is almost entirely true. — The variable C in the above content refers to circulation per 15-or-more-year-old inhabitant. — The series describing the number of papers distributed by mail, as published in *Industrial Statistics* is erroneous. It has been corrected by *Lehtien postimaksukomitea* (*Lehtien postimaksukomitean mietintö* 1959, p. 14). The corrected series was used in this study. The 1959 figure was obtained from *Posti- ja lennätinhallitus* (*Board of Posts and Telegraph*), and the number of papers transported by air by the publishers themselves was added to this figure.[1]

The income level of the newspapers' customers (subscribers, advertisers, etc.) is represented by the real national product per 15-or-more-year-old inhabitant (Y_{t-1}). In Finland this is substantially the same as real national income, the only notable difference being that real national income does not include the war reparations paid to the Soviet Union. It is desirable to include the variable Y_{t-1} in the analysis either separately or in the form of other variables closely correlated with it, because not only does it describe the income level of those who use the services offered by the papers, but it also includes influences not otherwise accounted for (*e.g.*, printed matter other than newspapers, ordered by the public).[2]

The price of newsprint (P_{t-1}) refers to the price of one kilo of newsprint. It has been obtained from *Paperikonttori*'s price adjustment sheets.

[1] Newspapers began their own air transport late in 1958.

[2] We also tried to include a variable in the analysis to represent changes in the distribution of income. The variable thus formed was the annual ratio to total income of the amount of income that should have been transferred from those with an income higher than the median to those with an income lower than the median in order to create an even distribution of income. In a period as short as that covered by this analysis, the above ratio seemed to vary so little that its use was restricted to tentative experiments (*cf.* BENTZEL 1953, p. 197). — We might have used family income here, but this was not available from official statistics. Another alternative would have been the real national income per married couple. It must be admitted, however, that both of these income concepts are rather illogical as explanations of fluctuations in the amount of newsprint sold due to changes in the consumption of advertising space. The real national product per 15-or-more-year-old inhabitant has the practical advantage of being easier to forecast since it does not call for a preview of the birth rate, etc.

Subscription rates (p_t) have been obtained from the annual means of the subscription rate index computed by *Lehtien postimaksukomitea*. This committee has estimated the index from a sample according to Laspeyres' index formula, using the estimated revenue from subscriptions in 1956 as weights (*Lehtien postimaksukomitean mietintö* 1959, p. 69).

The price of space for announcements and advertisements (π_{t-1}) is given in the form of the annual means of the advertising space rate index, computed by the above-mentioned committee. The method of computation resembles that used for the subscription rate index (*Lehtien postimaksukomitean mietintö* 1959, p. 69).

The revenue that the newspapers receive from advertisements ($R_{A_{t-1}}$) is described by the product of the variables A_{t-1} and π_{t-1}, and subscription revenue (R_{C_t}) by the product of variables C_t and p_t. As regards the content of R_{C_t}, the same remarks apply to it as for C_t , above.

The economic activity of the advertisers (S_{t-1}) is described by the volume index of retail and gross sales per 15-or-more-year-old inhabitant. The requisite volume index was computed by the *Central Statistical Office*.

The index of cost level (G_{t-1}) has three components: (1) editing costs, for which we used a substitute series describing the total salaries in the printing industries (printing offices proper, lithographic, photogravure and plate printing plants), deflated by the price index of the net national product (see Table III and Chapter 623, below); (2) total wages of the newspapers' wage earners, deflated by the price index of the net national product, published in *Teollisuustilastoa* (*Industrial Statistics of Finland*) since 1954; the figures prior to 1954 were computed by assuming that the wages of wage earners employed by the newspaper publishers formed 6.5 per cent of the total wages of all wage earners in the printing industries; in 1954—59 this proportion varied between 5.9 and 7.1 per cent; (3) cost to the newspapers of newsprint. This was obtained as a product of the quantities of newsprint sold each year and the corresponding unit prices (Table II, p. 73). The deflated series obtained from points (2) and (3) were added together to form a new series from which an index was computed. Another index was computed from the deflated series referred to in (1), above. In combining these two component indexes, that obtained from component (1) was weighted by 27 and that obtained from the sum of (2) and (3) was weighted by 57. These weights were obtained by reference to *Lehtien postimaksukomitean mietintö* (1959, p. 55) and they were largely the reason for the somewhat complicated computation procedure. The above report gives the ratio of editing costs (27 per cent) and printing and newsprint costs (57 per cent) to total costs (printing, editing, distribution and transport — not administration, maintenance, etc.) in 1956. To allow for changes in productivity, the ultimate cost index in this study was computed per kilo of newsprint bought each year. The use of constant weights here is inaccurate insomuch as the absolute amount of newsprint costs during the observation period increased roughly three-fold, while printing and editing costs merely doubled. Table III (p. 74) shows the different stages of computing the index.

The business cycle variable (B_t) was computed by dividing the observed volume index values of total exports in 1949—59 by the corresponding linear trend values. The trend values were obtained from the equation $X = 6.173\,t + 59.782$, in which X = volume index of exports and t = time (see Table II, p. 73).[1]

Preliminary experiments indicated that the above variables were probably adequate for the construction of the model.

622.2 Homogeneity of Observations

It may be questioned whether all the observed values of the variables used are free from the influence of exceptional restricting factors (factors that cannot be accounted for by a random variable). It was for this reason that the period of paper rationing was omitted from the data (see p. 20, above). The lagged observed values of some of the variables refer to 1948, but this was not considered to affect the results. On the other hand, the data for 1956 were formed under conditions that influence the different variables in different ways. For instance, the General Strike of that year affected the real national product much less, relatively speaking, than it affected the circulation of newspapers, and the consumption of advertising space and newsprint.

There seem to be three alternatives in dealing with a deviating observation: (a) to use the observation as such; (b) to adjust it by some logical method independent of the aims of the study; (c) to eliminate it. — Alternative (a) can be used if it is difficult to discover an objective method of adjustment and if it is desired not to lose a degree of freedom. It may also be justified in certain cases if the significance of the individual parameters is secondary in comparison with the forecasting properties of the model. Alternative (b) seems warranted in cases where it is desired not to lose a degree of freedom and where the aim is to implement a certain policy by single variables in a model free from multicollinearity. The »correct» estimates of the parameters are then more important than they are in a forecast model. Alternative (c) is justified where the loss of one degree of freedom is not prejudicial to the study. If the observation differs essentially from the others and no objective adjustment method exists, this alternative may the only feasible one (see PÖYHÖNEN 1955, pp. 34—35).

In this study the observed values for 1956 of the variables have been included

[1] Other alternatives for calculating a business cycle variable were considered (foreign currency reserves, value of exports, level of employment, volume of imports, etc.). Some of these rested on inadequately logical foundation; others were either lacking in data or provided a poor fit.

in the analysis unchanged because this is not considered to affect comparison between the models. After one or more models that comply adequately with the advance requirements have been chosen, they will be tested against data in which the values of the variables observed in 1956 have been adjusted in the manner indicated in the context. In this way, it may be possible to avoid adjusting variables in cases in which we should have to compute the seasonal index for the adjustment.

623 Deflating the Time Series

Since the aim is to construct a model valid for forecasting, it may be most practical to express all the monetary variables in terms of the purchasing power of the currency at a certain base period. This would eliminate the necessity either of including a separate inflation variable in the model or of assuming that the change in the purchasing power is constant. The selection of a deflator, however, is difficult and, to some extent, arbitrary. It was first intended in this study to describe the income level by the net national product. There seemed, however, to be no time series suitable for deflating the net national product, price of newsprint, and subscription and advertising rates. It was decided that the deflator had to be uniform for all the monetary variables. The final procedure adopted was to replace the net national product by the real national product in describing the income variable and to deflate the other monetary variables by the price index of the net national product.[1] The latter can be calculated from official statistics by dividing the net national product by its volume.

The difference between the effect of alternative deflators is not easy to describe. One logical conclusion, however, can be drawn — *i.e.* that, contrary to the wholesale price index and certain other indexes, the price index of the net national product takes account of the substitution effects because it is calculated with the aid of variable quantity weights. Although newsprint can hardly be replaced by other types of paper, other mass communication media may curtail the consumption of both news and advertising space. Similarly, the price index of the net national product takes account of newsprint that the public receives free of charge in »newspapers» and other printed matter circulated by industrial, political and intellectual organizations, and large co-operatives, though the weights used are not necessarily accurate.

624 Method of Computation

Several methods of multiple regression analysis have been developed with a view to rationalizing the work of computation. The most important practical consideration has been to reduce costs by developing as simple but informative

[1] The use of this index was suggested by Mr. JUSSI LINNAMO (Lic. Soc. Sc.).

a method as possible of computing the regression coefficients and their standard errors, and the coefficient of multiple correlation. Of the methods known to date that can be adapted to electric calculators, the most advantageous is that elaborated by TÖRNQVIST (1957, pp. 219—226; *cf. ibid* 1958, pp. 150—159; LARNA 1959, pp. 160—166). It is based on the method of least squares and offers, *inter alia*, the following advantages:

— One explanatory variable can be added to the analysis at a time.

— One or more variables can be eliminated from the analysis without being necessary to reject computations performed before the inclusion of the variable to be eliminated.

— The method gives an idea of the collinearity (see p. 33, below) of the explanatory variables; the more collinear these are, the more the main diagonal elements and, eventually, the other elements of the inverse matrix tend to grow (the matrix »explodes»).

— It offers double facilities for checking both the regression coefficients and the residnal square sums.

— Bunch map analysis describing the multicollinearity in the model, can be made on the basis of the inverse matrix.

— Observations can be added to or eliminated from the analysis with no need to recompute the moment matrix.

— The last stage of the inverse matrix provides the elements for computing the standard errors of individual estimates (under certain conditions, also those of forecasts) (see p. 57, below).

In this study, TÖRNQVIST's method was used with the aid of an electric desk calculator in numerous preliminary experiments and, later on, in supplementary experiments. To maintain reasonable numerical accuracy in the estimates, the decimal point in the observed values was shifted to ensure adequate decimal accuracy when checking the elements of the moment matrix computed from the logarithms of the observations. This was so successful that as late as the sixth inversion stage of the moment matrix no additional adjustment was required. All the checks then indicated an accuracy to 6—8 decimal places.

Some of the calculations were done on the electronic computer of The Finnish Cable Works Co. Ltd,[1] for which TÖRNQVIST's method had been programmed. The same programme supplied the residual term in addition to the customary information obtained from TÖRNQVIST's method.

625 Multicollinearity

Figures 1—3 (pp. 21—24, above) indicate which variables can be included in the model in order to obtain a logical result. On the other hand, it is impossible to construct a model with the most desirable econometric properties without experimentation. These properties are usually evaluated by means of the multiple coefficient of correlation and the significance level of the regression coefficients.

[1] The computations were performed by Mr. SEPPO MUSTONEN (M.A.).

A logically acceptable model with a seemingly high explanatory power, however, may sometimes have to be rejected because of multicollinearity.

A model or a set of data is said to contain multicollinearity when linear relationships exist not only between the variable to be explained and each explanatory variable, but also between the explanatory variables. Such is the case, for example, where variables distinctly different from each other in theory vary to either side of their linear trend so little that it is impossible to determine the influence of each of them on the variable to be explained. The same phenomenon may also be present in cross-section data, though it is not referred to then as a trend. The explanatory variables may then receive arbitrary coefficients. The coefficient of multiple correlation is liable to become very high. On the other hand, the significance of the regression coefficients does not rise in proportion to the reduction of the residual square sum. In the inverse matrix obtained by TÖRNQVIST's method this takes the form of an increase in the main diagonal elements used as coefficients in computing standard errors.

Multicollinearity can be described by means of »bunch-map analysis» (FRISCH 1934). This method has not been used here. It is a laborious method and not considered worth while, as it does not radically improve the picture obtained from the correlation matrix of all variables and from TÖRNQVIST's inverse matrix (see VALAVANIS 1959, p. 149).

In general, multicollinearity is regarded as a negative quality. It is difficult to deduce from a model in which it is present the net influence of one explanatory variable on the variable to be explained. This may be the case where it is desired to use a variable as a means of implementing a certain policy. If, however, the purpose of the model is merely to serve as a basis for forecasting single points by means of new observed values of all explanatory variables, multicollinearity is detrimental only where it is apparent but not real. Here it is said to be apparent if it is due mainly to a linear trend common to all the explanatory variables. If this trend is removed, either by trend-removal methods proper or by transforming the observations into differences of a certain order, multicollinearity is either drastically reduced or else it remains. In this study, the coefficients obtained from the original data were compared with those obtained from first logarithmic differences. The better these corresponded to each other, the more it seemed that the multicollinearity in the model was a real property and not an apparent one due to a linear trend in all the variables. In cases like this it must be borne in mind that the model can only be used with reservations for any purpose except forecasting new values for the variable to be explained by new observed values of all explanatory variables. For example, there can be no justification for computing partial regression coefficients. Similarly, the parameter estimates obtained cannot be regarded as parameters of action of the economy under consideration.

626 Selection of Variables for Experiments

As stated in Chapter 614 (pp. 20—25, above), the variables adopted to explain newsprint sales can be chosen by two different methods. In one, the factors affecting decision making in newsprint purchases are considered with a simultaneous analysis of the lags. In the other, it is endeavoured to divide newsprint consumption into components with a view to analysing each of these separately. This model, from which it is possible to achieve simple applications by combining the equations and using substitute series is presented on pages 24—25, above. Thus the model implicit in equations (2) — (8) can be applied in the form of equations (10) — (12):

(10) $C = 10^{\alpha_1} H^{\alpha_2} Y^{\alpha_3} K^{\alpha_4} p^{\alpha_5}$;

(11) $Z = 10^{\beta_1} B^{\beta_2} T^{\beta_4} \pi^{\beta_5} 10^{\beta_6 t}$;

(12) $Q = CZ$;

in which

C = circulation
H = population
Y = real income per inhabitant
K = level of knowledge
p = average subscription rate for all newspapers
Z = consumption of news and advertising space
B = business cycle variable
t = time
T = volume of industry, house building and trade
π = rate for advertising space
Q = consumption of newsprint
$\alpha_1 \ldots \alpha_5, \beta_1 \ldots \beta_5$ are parameters.

The following linear equation is obtained by substituting the values of C and Z from equations (10) and (11) in equation (12), taking the logarithms of both sides of the resultant equation and adding the random variable, and expressing Q, Y and T per inhabitant.

(13) $\log Q = \gamma_1 + \gamma_2 \log Y + \gamma_3 \log K + \gamma_4 \log p + \gamma_5 \log B + \gamma_6 \log T + \gamma_7 \log \pi + \gamma_8 t + u = \log Q^* + u$ [1],

in which $\gamma_1 \ldots \gamma_8$ are parameters and u is a random variable.

Owing to multicollinearity and the small number of degrees of freedom, several of the variables in equation (13) must be eliminated when applying the model. If Q is described by the newsprint sales series, the lags between the variables

[1] $\log Q^*$ = the value of $\log Q$ calculated from the model.

Table 1. Correlation matrix of the variables used in the analyses performed with the aid of non-transformed [1] observations. 1949—59.

	log Q_t	log Y_{t-1}	log A_{t-1}	log C_t	log $R_{A_{t-1}}$	log R_{C_t}	log P_{t-1}	log p_t	log π_{t-1}	log G_{t-1}	log S_{t-1}	log B_t
log Q_t	1.000	.947	.965	.909	.955	.865	.777	.717	.556	.070	.914	.041
log Y_{t-1}		1.000	.968	.903	.932	.867	.602	.723	.449	–.079	.974	–.103
log A_{t-1}			1.000	.861	.981	.855	.663	.734	.538	–.122	.943	.009
log C_t				1.000	.795	.860	.629	.650	.249	.022	.922	–.012
log $R_{A_{t-1}}$					1.000	.832	.689	.742	.692	–.130	.896	.019
log R_{C_t}						1.000	.445	.946	.428	–.057	.816	–.163
log P_{t-1}							1.000	.270	.518	.525	.594	.158
log p_t								1.000	.486	–.217	.635	–.238
log π_{t-1}									1.000	–.108	.386	.054
log G_{t-1}										1.000	–.066	–.074
log S_{t-1}											1.000	–.066
log B_t												1.000

Source: Table IV on p. 75.

in equation (13) can be read or inferred from Figures 2 and 3 (pp. 23 and 24, above).

The subscription price of newspapers can be included in the model either separately or by using it to deflate Y. Equation (13), however, does not include the price of newsprint, which cannot be taken into account in such a development of the model. It would seem to be essential to include it in a model that is to take the seller's decisions into account. It may be more appropriate to construct a forecast model starting from the factors affecting the decisions of the newspaper publishers in purchasing newsprint. The correlation matrix of all the variables selected (Table 1) can be used as an aid.

627 Testing of Hypotheses

The logicality of the economic theory underlying the model is ensured when specifying the model, but its applicability to a given set of observations must be examined by means of statistical tests. For this purpose, we require certain *a priori* criteria on the basis of which the model can either be accepted or rejected.

The »explanatory power» of the model is described by the coefficient of multiple correlation which refers to the correlation between the logarithms of newsprint sales (per 15-or-more-year-old inhabitant) calculated according to the model and those of the sales observed (*cf.* Törnqvist 1957, pp. 219—226):

[1] »Non-transformed» observations refer to observations from which the first differences have not been computed.

$$(14)\ R = \sqrt{1 - \frac{q_r^{(n)}}{q_r^{(1)}} \cdot \frac{f^{(1)}}{f^{(n)}}},$$

in which R = the coefficient of multiple correlation,
$q_r^{(1)}$ = the residual sum of squares at the first inversion stage of the moment matrix (based on TÖRNQVIST's method),
$q_r^{(n)}$ = the residual sum of squares at the last inversion stage of the moment matrix,
$f^{(1)}$ = the number of degrees of freedom at the first inversion stage,
$f_r^{(n)}$ = the number of degrees of freedom at the last inversion stage.

The significance of the »explanation» obtained from the model can be tested by the variance ratio test (the F-test), by comparing the variance explained by the model with the residual variance. In this study R^2 (the relative variance explained) is so large that no F-test is necessary.

The significance of the parameter estimates can be evaluated by applying the t-test to determine whether they deviate significantly from zero. The following formula is used for this purpose:

$$(15)\ t = \frac{\alpha_i}{\varepsilon_{\alpha_i}},$$

in which t = the test variable,
α_i = the estimate of the parameter being tested,
ε_{α_i} = the standard error of α_i.

Since the t-test, which is based on Student's distribution, presupposes that the explanatory variables are not intercorrelated, the t-test is used here (the data being multicollinear) as a conventional norm in comparing the properties of the different models with each other. In most instances t-values are not explicitly indicated because the standard errors in the parameter estimates obtained from different models are comparable as such (provided the number of degrees of freedom is the same). Throughout this study, however, only parameter estimates with a formal risk level of less than 5—10 per cent have been accepted.

628 Models Based on the Economy of Newspapers[1]

628.1 Value Series as Explanatory Variables

It can be assumed that newspapers anticipate their future revenue and expenditure partly in the light of previous experience, and partly on the basis of more precise information obtained in some other way. The publishers are obliged to ascertain their expectations when they budget their revenue and expenditure for the following year. Thus they may anticipate that the revenue from advertisements will develop in the same way as it has in previous years, while the bulk

[1] A synopsis of the estimation statistics for the models tried out in this study is given in Tables X—XIII, pp. 80—82.

of the revenue from subscriptions for a given year is known at the end of the preceding year. As regards the most important variable cost, that of newsprint, they may have to be satisfied with estimates based on experience. In a statistical analysis, however, the total cost of newsprint must be replaced by the unit price of newsprint because it is hardly feasible to explain variations in some quantity by variations in a revenue in which this quantity is the multiplier. After forming an idea of future revenue and expenditure, a newspaper is in a position to influence these and thus modify its annual turnover. This leads us to the same process of reasoning in making decisions regarding the purchase of newsprint as is shown in Figures 1—3 (pp. 21—24, above).

Value series here refer to time series composed of annual revenue from advertising and subscriptions deflated by the price index of the net national product (see p. 29, above). To start with, we tried out the model

$$(16)\ \log Q_t = \alpha_1 + \alpha_2 \log R_{A_{t-1}} + \alpha_3 \log R_{C_t} + \alpha_4 \log P_{t-1} + u_t,$$

in which $\alpha_1 \ldots \alpha_4$ are statistical parameters, and the variables are expressed in terms of index numbers (1949 = 1.00),

Q_t = sales of newsprint in kilos per inhabitant (15 or more years of age) during the year t
$R_{A_{t-1}}$ = advertising revenue per inhabitant[1] during the year t-1, deflated by the price index of the net national product
R_{C_t} = revenue from subscription per inhabitant in the year t, deflated by the price index of the net national product
P_{t-1} = price of newsprint in the year t-1, deflated by the price index of the net national product
u_t = the random variable.

The business cycle variable B_t (see p. 30, above) was also tried out in this model, but it did not improve the explanation significantly. It was not considered worth while in this equation to study the properties of the residual term. Instead, a new model was introduced in which the price of newsprint was replaced by the variable G_{t-1} describing the cost level in newspaper publishing (see p. 29, above). This was considered justifiable mainly because the cost of newsprint comprises only a small proportion of the total cost of publishing a newspaper (see pp. 21 and 22, above). After the substitution of G_{t-1} in the model, the specification of the model is easy to understand since the consumption of newsprint is explained there in terms of an input-output analysis. The parameter estimates (α_i) and their standard errors, the standard errors of estimates (s) from the models and the coefficients of multiple correlation (R) for models (16) and

$$(17)\ \log Q_t = \alpha_1 + \alpha_2 \log R_{A_{t-1}} + \alpha_3 \log R_{C_t} + \alpha_4 \log G_{t-1} + u_t,$$

are given in the following tabulation.

[1] »Inhabitant» throughout this study refers only to those 15 years of age or over.

Explanatory variable	Regression coefficient Model (16)	(17)	α_i
1	.102 ± .201	.081 ± .022	α_1
$\log R_{A_{t-1}}$	.420 ± .132	.774 ± .129	α_2
$\log R_{C_t}$	.551 ± .174	.302 ± .208	α_3
$\log P_{t-1}$	.855 ± .235	·	α_4
$\log G_{t-1}$	·	.896 ± .373	α_4
R	.982	.972	
s	.0228	.0286	

The business cycle variable (B_t) did not improve the explanatory power of model (17) significantly. The experiment with the value series seems to show that, taken alone, they hardly enable a satisfactory forecast model to be formulated.

628.2 Value and Volume Series as Explanatory Variables

Owing to the multicollinearity of the variables, it is difficult to decide without experimenting which combination of them provides the best application of the model based on the economy of the newspaper publishers. One of the value series, the subscription revenue (R_{C_t}) of the newspapers, was therefore replaced by the corresponding volume series (C_t), and the following model was obtained:

$$(18)\ \log Q_t = \alpha_1 + \alpha_2 \log R_{A_{t-1}} + \alpha_3 \log C_t + \alpha_4 \log P_{t-1} + u_t.$$

The business cycle variable, B_t, reduced the residual sum of the squares so little that the coefficient of multiple correlation decreased when it was inserted in the model. The coefficient of multiple correlation was raised considerably by the substitution of C_t for R_{C_t}. Similarly, the relative standard errors of $\alpha_2 \ldots \alpha_4$ were reduced.

The following model was also tried out:

$$(19)\ \log Q_t = \alpha_1 + \alpha_2 \log A_{t-1} + \alpha_3 \log R_{C_t} + \alpha_4 \log P_{t-1} + u_t,$$

in which $R_{A_{t-1}}$ in equation (17) was replaced by the corresponding volume index for advertising (*i.e.* advertising space consumed), A_{t-1}. The tabulation below gives the parameter estimates and other estimation statistics for models (18) and (19).

There is no major difference between the explanations given by models (18) and (19) and we shall not test its significance; the relevance of this test will depend on the short term properties of these models, which will be tried out later (pp. 42—43, below). At this stage, a comparison of Figures 4 and 5 will suffice to illustrate the differences between the two models.

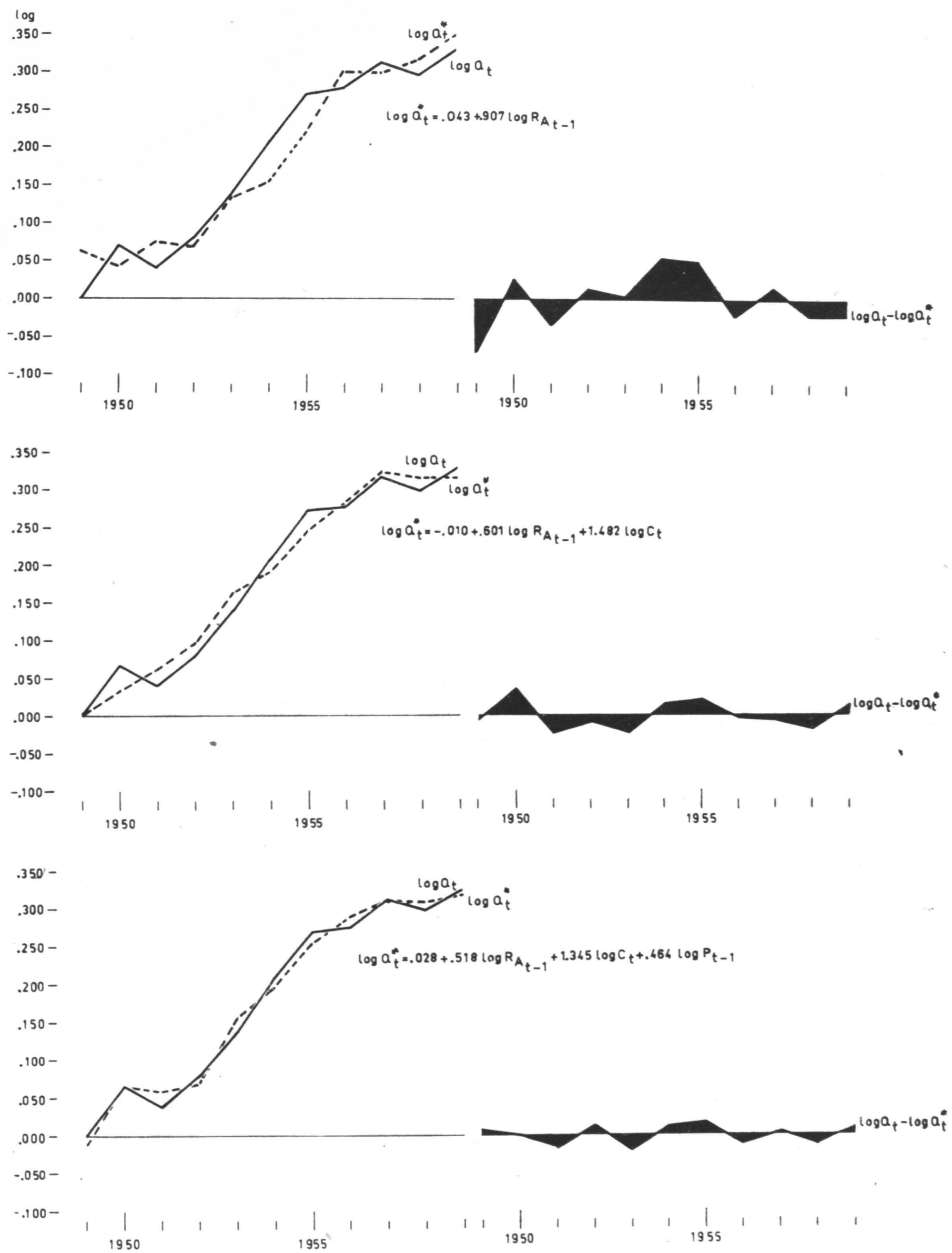


Figure 4. The logarithm of newsprint sales in Finland in 1949—59 (i) as observed ($\log Q_t$) and (ii) as calculated from Model (18) ($\log Q_t$*), and the corresponding residual term ($\log Q_t$ — $\log Q_t$*) with the increase of one variable at a time in the analysis.

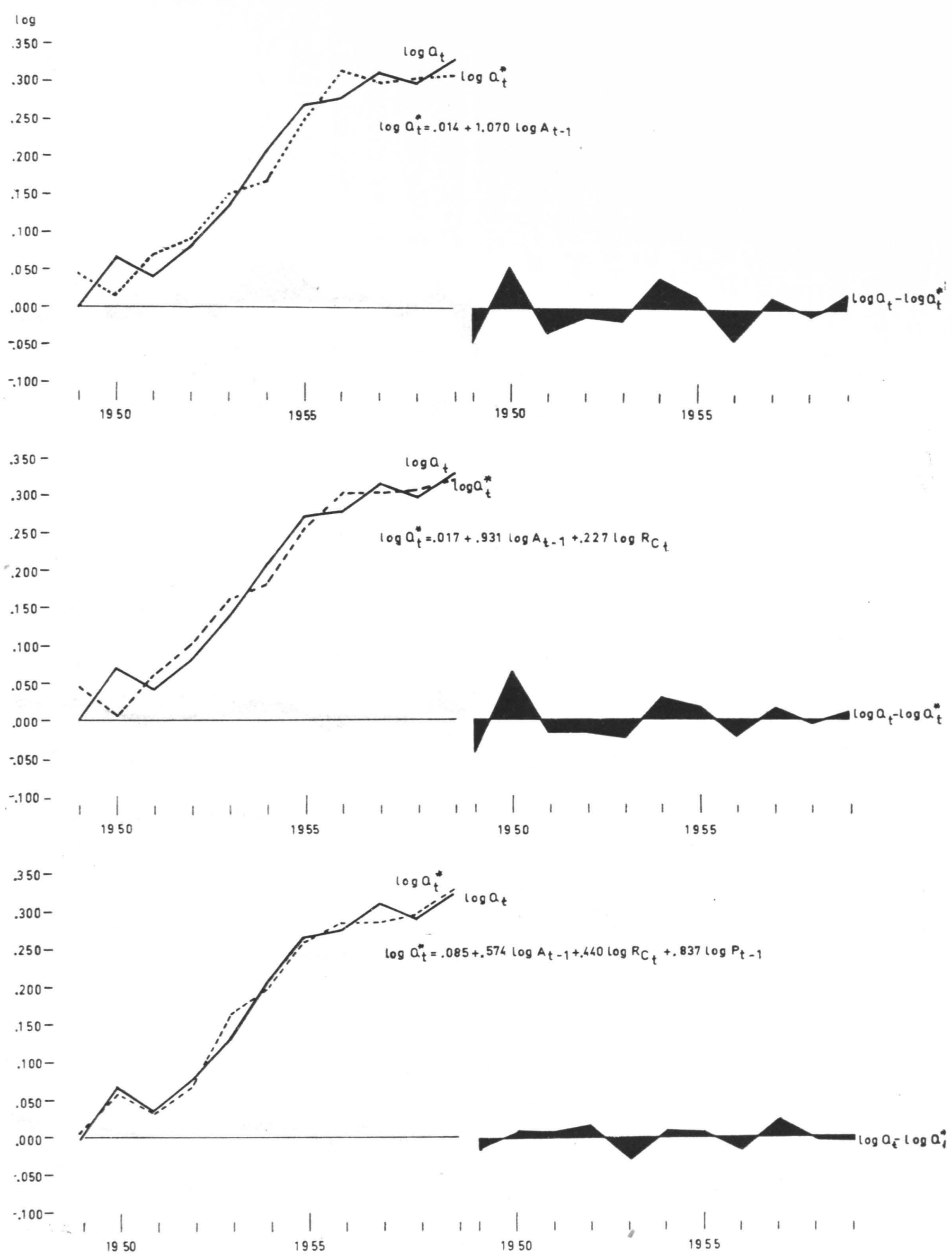


Figure 5. The logarithm of newsprint sales in Finland in 1949—59 (i) as observed (log Q_t) and (ii) as calculated from Model (19) (log Q_t*), and the corresponding residual term (log Q_t — log Q_t*) with the increase of one variable at a time in the analysis.

Explanatory variable	Regression coefficient Model (18)	(19)	α_i
1	.028 ± .018	.085 ± .017	α_1
log $R_{A_{t-1}}$	.518 ± .070	•	α_2
log A_{t-1}	•	.574 ± .123	α_2
log C_t	1.345 ± .253	•	α_3
log R_{C_t}	•	.440 ± .143	α_3
log P_{t-1}	.464 ± .157	.837 ± .174	α_4
R	.991	.990	
s	.0158	.0175	

628.3 Volume Series as Explanatory Variables

With a view to simplifying the analysis we can also try substituting simple volume series for the advertising and subscription revenues. This may seem naïve seeing that the consumption index of advertising space and the circulation index are, if measured in a certain way, quantity components of newsprint consumption. This is not, however, an insuperable barrier in specifying a forecast model. Explanation here is of secondary importance as compared with description, though it is not necessarily to be avoided. The consumption index of advertising space, A_{t-1}, measures the average utilization of page area of one issue, while C_t measures its circulation. The estimates of the parameters and their standard errors for the model

(20) $\log Q_t = \alpha_1 + \alpha_2 \log A_{t-1} + \alpha_3 \log C_t + \alpha_4 \log P_{t-1} + u_t$

were as follows:

$$\alpha_1 = .036 \pm .022$$
$$\alpha_2 = .670 \pm .118$$
$$\alpha_3 = .916 \pm .375$$
$$\alpha_4 = .599 \pm .189$$

The coefficient of multiple correlation $R = .987$
The standard error of estimate $s = .0198$

The substitution of the business cycle variable, B_t, in the model proved to be unnecessary, for it did not significantly reduce the residual sum of the squares.

628.4 Conclusions Regarding Models Based on the Economy of Newspapers

In the models based on the economy of the newspapers, factors affecting the decision making in newsprint purchases that are rational in terms of either the achievement of the turnover or the general principles of input-output analysis are chosen to »explain» the annual amounts of newsprint bought. If forecasting is the only purpose of the model, there is no objection to using these models to describe the economic structure of the newspapers. Model (18) is adequate in description; its coefficient of multiple correlation is .991 despite the small number of degrees of freedom available. As Figures 1—3 (pp. 21—24, above) show, the explanatory variables are predetermined in respect of Q_t, even though the lags of C_t and R_{C_t} are so small that no explicit account can be taken of them. There is, however, some reason to believe that a better result can be achieved with a model in which the decisions of the subscribers and advertisers are stated more expressly. This possibility is indicated by the fact, for example, that newspapers cannot maintain large stocks of newsprint, but can, for this very reason evaluate the decisions they made at the end of the preceding year in the light of their customers' decisions. To allow for simultaneous influences, the non-lagged price of newsprint, P_t, was tried out in the model in addition to the lagged price, P_{t-1}. In no case did it prove significant.

The constant α_1 may perhaps describe structural changes if it varies significantly when solved for consecutive periods from an equation consisting merely of variables describing the economic behaviour of the newspapers' customers [1] (*cf.* STONE and ROWE 1955, p. 9). From the point of view of forecasting, however, the additional information α_1 may offer is by no means indispensable.

It is difficult to gain an insight into the properties of the models, if their parameters have been estimated merely from data containing the trend. The trend in all the variables selected may be caused by an outside factor. Hence the »explanation» obtained from the model may be mainly arbitrary. On the other hand, it is desirable to obtain an explanation of the trend from the model, even if such a trend common to all variables is the reason for their multicollinearity (see p. 33, above). True, it is difficult to ascertain when a common linear trend is caused by an outside factor. In this study, it was decided to compare the parameters and correlation coefficients estimated from the original data with those estimated from transformed data. The better these correspond to each other, the less reason there is to assume that the explanation obtained is due to a linear trend common to all the variables. However, this does not necessarily eliminate multicollinearity. The variables may be correlated even after a linear trend has been removed. Such is the case sometimes in business cycles, for instance.

Several methods have been developed with a view to reducing the mutual dependence of consecutive observations (*cf.* TINTNER 1952, pp. 301—329). In this study, we computed the first logarithmic differences [2] of the observations,

[1] A newspaper's »customers» refer in this study to subscribers and advertisers together (plus job-printing, etc. clients, if any).

[2] The correlation matrix of these is set out in Table 2. Logarithmic differences were obtained by taking the logarithms of the successive ratios of the observations [log $\Delta X = \log (X_t / X_{t-1})$].

Table 2. Correlation matrix of the variables given in Table 1, as computed from the first logarithmic differences of the observations. 1949—59.

	log ΔQ_t	log ΔY_{t-1}	log ΔA_{t-1}	log ΔC_t	log $\Delta R_{A_{t-1}}$	log ΔR_{C_t}	log ΔP_{t-1}	log Δp_t	log $\Delta \pi_{t-1}$	log ΔG_{t-1}	log ΔS_{t-1}	log ΔB_t
log ΔQ_t	1.000	.038	–.153	.530	–.123	–.067	.675	–.266	–.001	.639	–.085	.035
log ΔY_{t-1}		1.000	.393	.091	.089	.143	–.310	.105	–.667	–.193	.573	–.500
log ΔA_{t-1}			1.000	–.354	.898	–.057	–.263	.067	–.221	–.494	.551	.085
log ΔC_t				1.000	–.516	.244	.177	–.123	–.430	.270	.199	–.032
log $\Delta R_{A_{t-1}}$ ..					1.000	–.219	–.072	–.039	.226	–.368	.382	.160
log ΔR_{C_t}						1.000	–.578	.932	–.359	–.264	–.208	–.268
log ΔP_{t-1}							1.000	–.650	.381	.874	–.112	–.024
log Δp_t								1.000	–.207	–.370	–.295	–.266
log $\Delta \pi_{t-1}$									1.000	.264	–.362	.193
log ΔG_{t-1}										1.000	–.143	–.256
log ΔS_{t-1}											1.000	–.116
log ΔB_t												1.000

Source: Table VIII on p. 78

and carried out the same analyses with these as with the original data. This gave the following equation corresponding to equation (16):

$$\text{(16a)} \quad \log \Delta Q_t = \alpha_1 + \alpha_2 \log \Delta R_{A_{t-1}} + \alpha_3 \log \Delta R_{C_t} + \alpha_4 \log \Delta P_{t-1} + u_t.$$

Equations (18a), (19a) and (20a) were formed by the same analogy. The following tabulation shows the estimates of the parameters and other estimation statistics as computed for these models.

Explanatory variable	Regression coefficient Model (16a)	(18a)	(19a)	(20a)	α_i
1	.016 ± .145	.013 ± .015	.013 ± .013	.013 ± .013	α_1
log $\Delta R_{A_{t-1}}$	.056 ± .253	.179 ± .265	·	·	α_2
log ΔA_{t-1}	·	·	.139 ± .254	.177 ± .249	α_2
log ΔR_{C_t}	.391 ± .252	·	.409 ± .242	·	α_3
log ΔC_t	·	1.199 ± .651	·	1.109 ± .589	α_3
log ΔP_{t-1}	.663 ± .218	.406 ± .166	.695 ± .222	.435 ± .169	α_4
R	.651	.695	.669	.698	
s	.0262	.0248	.0257	.0247	

In general, the estimates obtained differ considerably from the corresponding estimates computed from original data. Their relative standard errors are large except in the case of α_4, which seems to deviate almost significantly from zero in all the equations ($t > 2.45$, 6 degrees of freedom, assuming that the testing conditions are present). The coefficients of multiple correlation are rather small; from this it can be concluded that models based on the economy of the newspapers give a poor »explanation» ($R^2 < .5$) of the short term variation of newsprint sales. Neither is there justification for assuming that models (16) — (20) »explain» long-term variation any better. The apparent explanation obtained may be due to a common trend.

629 Other Models

629.1 Models Derived from the Economy of the Newspapers

Model (20) (p. 41, above) can be developed so that the variable describing the consumption of advertising space, A_{t-1}, is replaced by the variable S_{t-1} — volume of wholesale and retail sales (see p. 29, above), which might be expected to explain the variation in A_{t-1}. It is perhaps reasonable to assume that most of the volume of advertising varies with changes in the volume of sales, though a minor part is more or less constant. In the model thus formed,

$$(21)\ \log Q_t = \alpha_1 + \alpha_2 \log S_{t-1} + \alpha_3 \log C_t + \alpha_4 \log P_{t-1} + u_t,$$

estimates of the parameters were as follows:

$$\alpha_1 = .070 \pm .045$$
$$\alpha_2 = .648 \pm .355$$
$$\alpha_3 = .972 \pm .979$$
$$\alpha_4 = .897 \pm .352$$

The coefficient of multiple correlation $R = .948$
The standard error of estimate $s = .0385$

Analysis based on the first logarithmic differences gave the parameters of the model,

$$(21a)\ \log \Delta Q_t = \alpha_1 + \alpha_2 \log \Delta S_{t-1} + \alpha_3 \log \Delta C_t + \alpha_4 \log \Delta P_{t-1} + u_t,$$

the following estimates and standard errors:

$$\alpha_1 = .021 \pm .105$$
$$\alpha_2 = .100 \pm .232$$
$$\alpha_3 = 1.030 \pm .586$$
$$\alpha_4 = .397 \pm .171$$

The coefficient of multiple correlation $R = .679$
The standard error of estimate $s = .0254$

The business cycle variable, B_t, did not prove to be significant in these models.

On the other hand, when model (16) was transformed by substituting S_{t-1} for $R_{A_{t-1}}$, the resultant model,

$$(22)\ \log Q_t = \alpha_1 + \alpha_2 \log S_{t-1} + \alpha_3 \log R_{C_t} + \alpha_4 \log P_{t-1} + u_t,$$

gave the following estimates for the parameters, together with their standard errors:

$$\alpha_1 = .109 \pm .020$$
$$\alpha_2 = .487 \pm .162$$
$$\alpha_3 = .621 \pm .164$$
$$\alpha_4 = 1.054 \pm .208$$

The coefficient of multiple correlation $R = .981$
The standard error of estimate $s = .0236$

By analogy, a model based on the first logarithmic differences was formed:

$$(22a)\ \log \Delta Q_t = \alpha_1 + \alpha_2 \log \Delta S_{t-1} + \alpha_3 \log \Delta R_{C_t} + \alpha_4 \log \Delta P_{t-1} + u_t.$$

The parameters of this model and their standard errors were estimated as follows:

$$\alpha_1 = .014 \pm .013$$
$$\alpha_2 = .129 \pm .244$$
$$\alpha_3 = .418 \pm .248$$
$$\alpha_4 = .648 \pm .216$$

The coefficient of multiple correlation $R = .667$
The standard error of estimate $s = .0257$

The estimates of the parameters of model (22), based on the original data, seem to deviate almost significantly from zero (with a probability of $\geqq$ 95 per cent). However, the substitution of the variable S_{t-1} for A_{t-1} or $R_{A_{t-1}}$ did not improve the models originally based on the economy of newspapers. These changes impaired the total explanation of the long-term variation. The standard errors of the parameter estimates were large, and the short term variation was explained inadequately.

629.2 Models Based on the Economy of the Newspapers' Customers

On the basis of Figures 1—3 (pp. 21—24, above), it might be justifiable to conclude that the effect of the decisions of the newspapers' customers — subscribers and advertisers — finds its way into variables R_{C_t} and $R_{A_{t-1}}$. This is true to a large extent, of course. On the other hand, the variables omitted from the applications may include influences that do not appear in the decision making of the subscribers and advertisers, but which do appear in newsprint consumption. Such influences are sales of single copies, printing of matter other than newspapers, printed matter distributed to the public free of charge, etc. It may be possible to account for some of these influences by deflating the monetary variables explaining newsprint consumption by the price index of the net national product (see p. 31, above). A more direct way of approaching this objective would be to include the real national product, Y_{t-1}, in the model as the variable describing the income level (see p. 28, above). Other variables representing the income level of either the newspapers or their customers can then be omitted. After preliminary experiments, the following model was obtained:

$$(23)\ \log Q_t = \alpha_1 + \alpha_2 \log Y_{t-1} + \alpha_3 \log p_t + \alpha_4 \log P_{t-1} + \alpha_5 \log B_t + u_t.$$

The values of the parameter estimates and their standard errors were as follows:

$$\alpha_1 = .108 \pm .016$$
$$\alpha_2 = 1.572 \pm .198$$
$$\alpha_3 = .510 \pm .142$$
$$\alpha_4 = .948 \pm .148$$
$$\alpha_5 = .500 \pm .215$$

The coefficient of multiple correlation $R = .992$
The standard error of estimate $s = .0155$

At this stage it is considered worth-while studying the autocorrelation of the residual term ($\log Q_t - \log Q_t^*$) so as to obtain some indication of either the absence of variables or the regularity of the residual term (whether systematic or random). If the residual term is autocorrelated, it may be possible to forecast its autocorrelated part in future time series, too, while the residual term thus created may serve as an estimate of the random variable, u_t. Here, only the first autocorrelation coefficient (the coefficient of one-year-lag correlation) was computed, since the number of degrees of freedom was small and the residual drawn (Fig. 6) did not suggest further experiments. The coefficient of autocorrelation ($r_{q_t q_{t-1}}$) was obtained from the following equation:

$$(24)\ r_{q_t q_{t-1}} = \frac{p}{s_{q_t} \cdot s_{q_{t-1}}},$$

in which

$$q = \log Q_t - \log Q_t^* \text{ (= observed logarithm — calculated logarithm)}$$

$$p = \frac{\Sigma (q_t q_{t-1})}{N} - \frac{\Sigma q_t}{N} \cdot \frac{\Sigma q_{t-1}}{N}$$

$$s_{q_t} = \sqrt{\frac{\Sigma q_t^2}{N} - \left(\frac{\Sigma q_t}{N}\right)^2}$$

$$s_{q_{t-1}} = \sqrt{\frac{\Sigma q_{t-1}^2}{N} - \left(\frac{\Sigma q_{t-1}}{N}\right)^2}$$

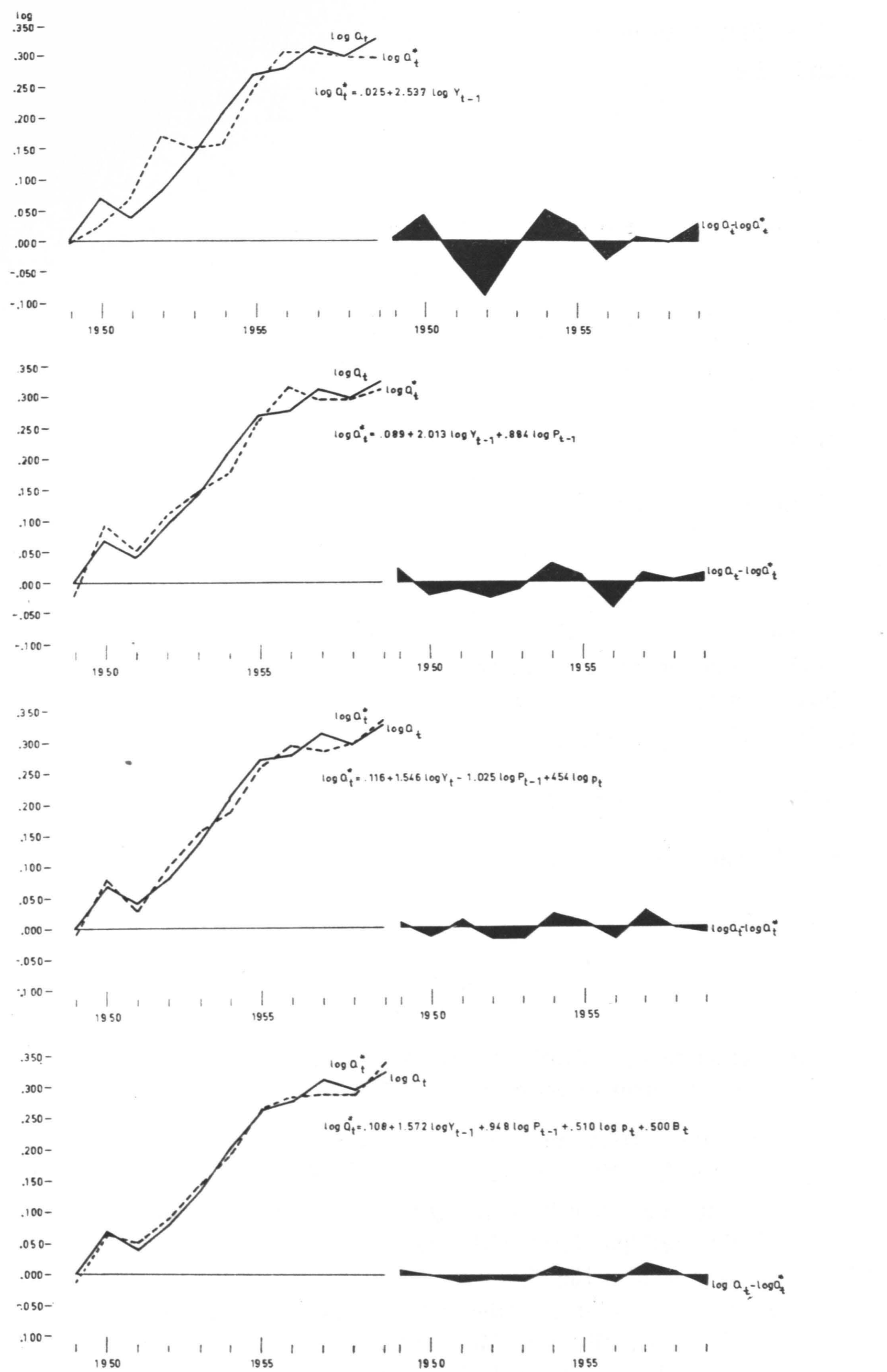


Figure 6. The logarithm of newsprint sales in Finland in 1949—59 (i) as observed ($\log Q_t$) and (ii) as calculated from Model (23) ($\log Q_t^*$), and the corresponding residual term ($\log Q_t - \log Q_t^*$) with the increase of one variable at a time in the analysis.

The first autocorrelation coefficient of model (23) was $r_{q_t q_{t-1}} = .163$. Its standard error($\varepsilon_{r_{q_t q_{t-1}}}$) was obtained from the formula

$$(25)\quad \varepsilon_{r_{q_t q_{t-1}}} = \frac{1 - r_{q_t q_{t-1}}{}^2}{\sqrt{N-2}}$$

Here, the value of $\varepsilon_{r_{q_t q_{t-1}}}$ came out to $\pm .344$, from which it can safely be concluded that the first autocorrelation coefficient of the residual term did not differ significantly from zero.

The specification of model (23) comes closer to an explanation than that of any of the models tried out previously in this study. The standard errors of its parameter estimates are small. α_5, the relative standard error of which is the largest, differs from zero with a probability of more than 90 per cent ($t = 2.33$, 6 degrees of freedom). Fig. 6 shows the development of the calculated values of log Q_t (log Q_t*) and the corresponding residual term with the insertion of one variable at a time into the analysis.

Further examination of model (23) reveals that it includes no variable which accounts for changes in subscription habits. During short periods such changes are hardly significant. Regardless of the length of the observation period, however, one or more »structural variable» could be included in the model. Thus, in model (23), p_t was replaced by R_{C_t}, which in fact means inserting C_t in the model. The model thus created,

$$(26)\ \log Q_t = \alpha_1 + \alpha_2 \log Y_{t-1} + \alpha_3 \log R_{C_t} + \alpha_4 \log P_{t-1} + \alpha_5 \log B_t + u_t,$$

gave the parameters and their standard errors the following estimates:

$$\alpha_1 = .091 \pm .013$$
$$\alpha_2 = 1.322 \pm .233$$
$$\alpha_3 = .470 \pm .120$$
$$\alpha_4 = .900 \pm .136$$
$$\alpha_5 = .449 \pm .200$$

The coefficient of multiple correlation $R = .993$
The standard error of estimate $s = .0145$

The first autocorrelation coefficient of the residual term of model (26) did not differ significantly from zero ($r_{q_t q_{t-1}} = .169 \pm .343$). The development of the newsprint sales quantities (log Q_t*) calculated from model (26), and of the corresponding residual term with the insertion of one variable at a time into the analysis is seen in Fig. 7.

Models (23) and (26) describe data containing both short and long term variation almost equally well. Model (23) is somewhat simpler to use, since no data on circulation are needed. The content of the variable R_{C_t} is not quite

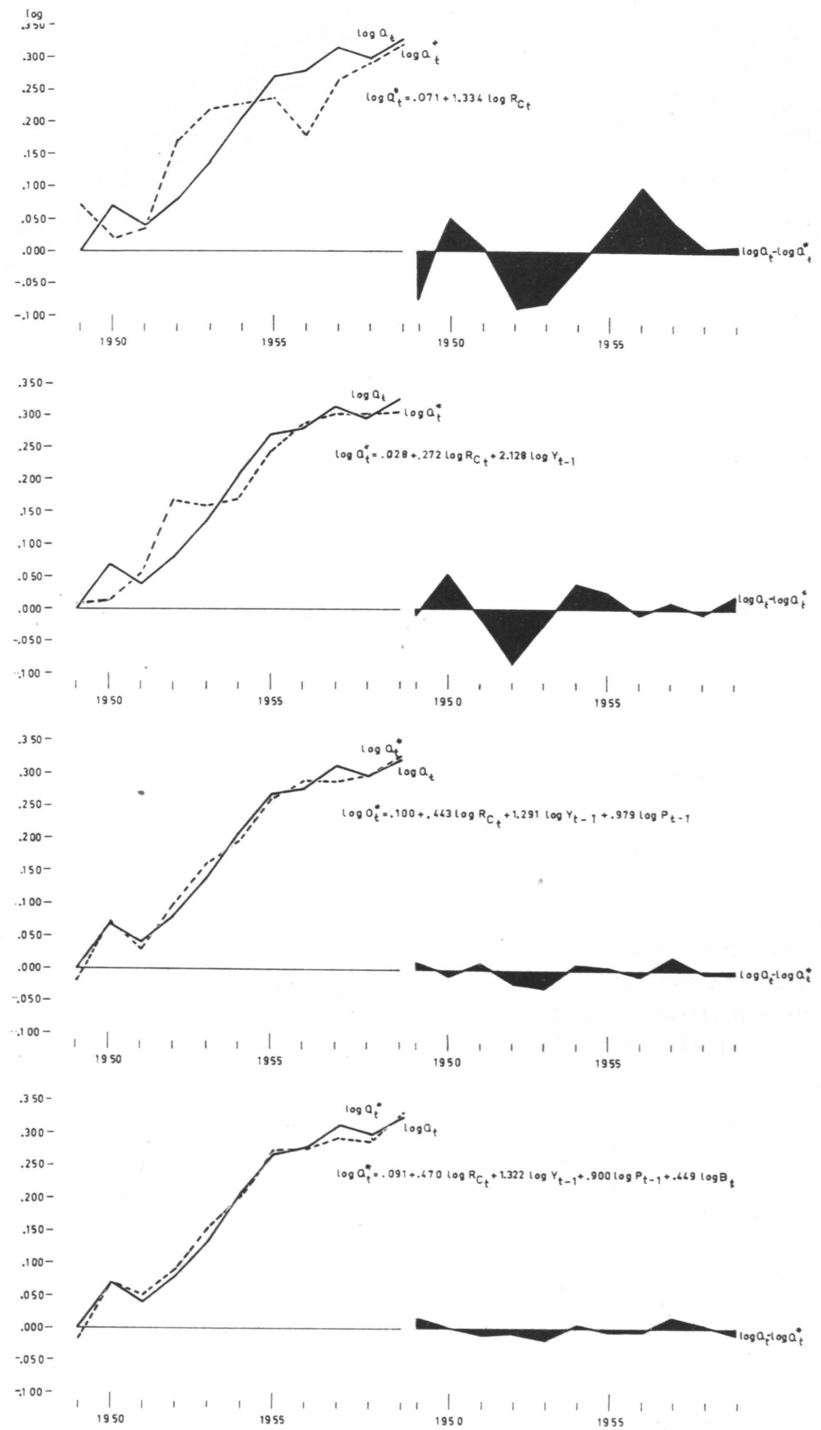


Figure 7. The logarithm of newsprint sales in Finland in 1949—59 (i) as observed (log Q_t) and (ii) as calculated from Model (26) (log Q_t*), and the corresponding residual term (log Q_t — log Q_t*) with the increase of one variable at a time in the analysis.

clear, owing to the nature of the data on circulation, and there is no certainty of its constancy in time (*cf.* p. 28, above).

The capacity of models (23) and (26) to describe short term variation were compared by means of an analysis based on the first logarithmic differences. On the analogy of model (23), model

$$(23a)\ \log \Delta Q_t = \alpha_1 + \alpha_2 \log \Delta Y_{t-1} + \alpha_3 \log \Delta p_t + \alpha_4 \log \Delta P_{t-1} + \alpha_5 \log \Delta B_t + u_t,$$

was formed and, on the analogy of model (26), model

$$(26a)\ \log \Delta Q_t = \alpha_1 + \alpha_2 \log \Delta Y_{t-1} + \alpha_3 \log \Delta R_{C_t} + \alpha_4 \log \Delta P_{t-1} + \alpha_5 \log \Delta B_t + u_t.$$

The following tabulation shows the parameter estimates, standard errors and the coefficients of multiple correlation obtained from these models.

Explanatory variable	Regression coefficient Model (23a)	(26a)	α_i
1	$-.001 \pm .013$	$-.005 \pm .009$	α_1
$\log \Delta Y_{t-1}$	$1.392 \pm .592$	$1.298 \pm .389$	α_2
$\log \Delta p_t$	$.550 \pm .258$	·	α_3
$\log \Delta R_{C_t}$	·	$.593 \pm .154$	α_3
$\log \Delta P_{t-1}$	$.924 \pm .234$	$.907 \pm .142$	α_4
$\log \Delta B_t$	$.499 \pm .247$	$.496 \pm .162$	α_5
R	.773	.898	
s	.0219	.0152	

The standard errors of $\alpha_1 \ldots \alpha_5$ are considerably larger in model (23a) than they are in model (26a). Model (23a) »explains» < 60 per cent of the variation of $\log Q_t$, while model (26a) »explains» > 80 per cent ($= R^2$). Thus, model (26) seems to be preferable to (23) in forecasting the values of Q_t one year in advance, provided that the observed value of R_{C_t} is known or can be forecast with reasonable accuracy. The changes in the parameter estimates for models (23) and (26), when transferred to models (23a) and (26a), are generally small and parallel. Fig. 8 shows the development of the values of $\log \Delta Q_t^*$ calculated from model (26a) when one variable at a time is inserted into the analysis.

Since models (23) and (26) seemed to be the most successful of all the models tried out, their parameters were also estimated from data in which the effect of the General Strike had been eliminated from the observations for 1956 (see p. 30, above). It was endeavoured to adjust the observations for 1956 in a manner independent of the objects of the study and the findings obtained hitherto.

The real national product index (Y_{t-1}) was enlarged by 1.5 per cent — based on LAURILA's (1956, pp. 38—39) estimate of the economic loss caused by the strike. The indexes of newspaper circulation (C_t) and newsprint sales (Q_t) were

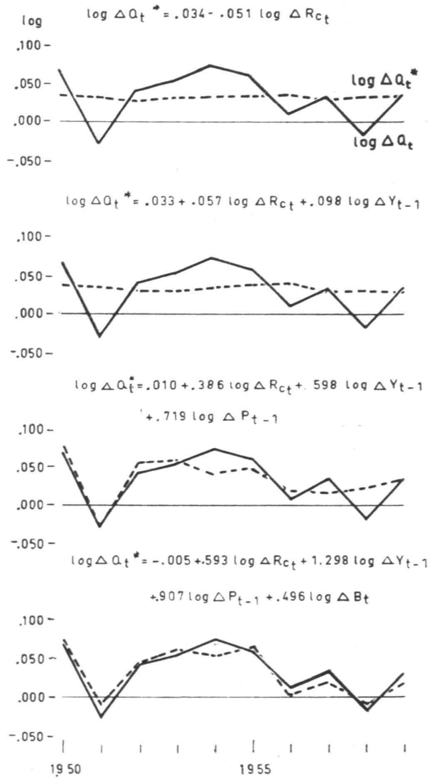


Figure 8. The year-to-year logarithmic differences of newsprint sales in Finland in 1949—59 (i) as observed ($\log \Delta Q_t$) and (ii) as calculated from Model (26a) ($\log \Delta Q_t^*$) with the increase of one variable at a time in the analysis.

raised by 8 per cent, which was considered roughly to correspond to the actual stoppage in terms of historical time. The influence of the strike on prices was not established, and no measures were taken to adjust the business cycle variable. The values of the adjusted variables are indicated in brackets in Table IV (p. 75).

The method of adjustment was very rough. The result offered by a more elaborate method hardly seemed to justify the effort required. As a whole, the General Strike lasted about three weeks, but in the printing industry it lasted a few days longer.

The »adjusted» parameter estimates and other estimation statistics for models (23) and (26) were as follows:

Explanatory variable	Regression coefficient Model (23) (with observations for 1956 adjusted)	(26)	α_i
1	$.105 \pm .014$	$.092 \pm .012$	α_1
$\log Y_{t-1}$	$1.715 \pm .171$	$1.439 \pm .239$	α_2
$\log p_t$	$.374 \pm .125$	·	α_3
$\log R_{C_t}$	·	$.396 \pm .129$	α_3
$\log P_{t-1}$	$.977 \pm .131$	$.927 \pm .126$	α_4
$\log B_t$	$.378 \pm .191$	$.380 \pm .188$	α_5
R	.994	.994	
s	.0138	.0136	

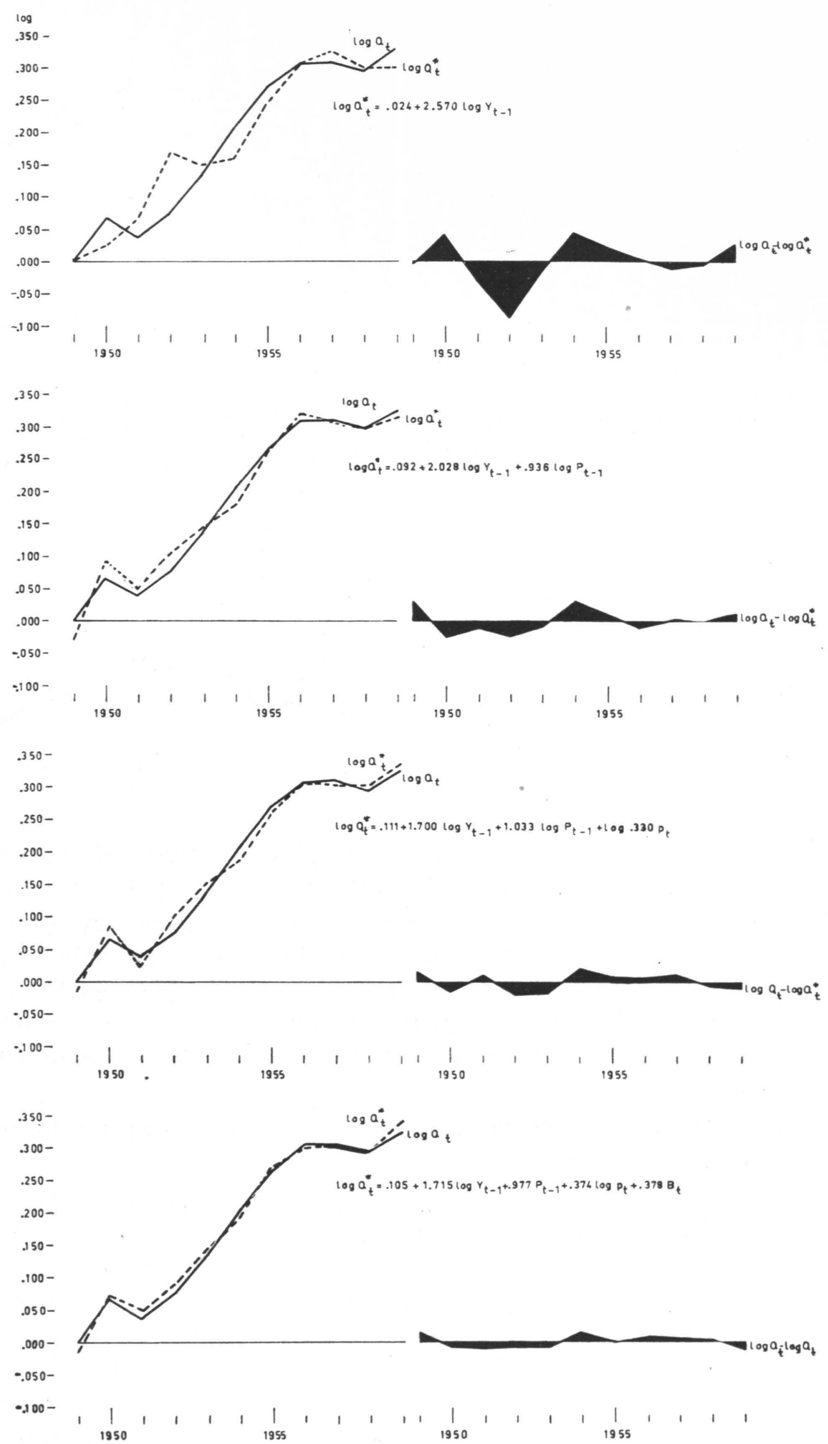


Figure 9. The logarithm of newsprint sales in Finland in 1949—59 (i) as observed (log Q_t) and (ii) as calculated from Model (23) (log Q_t*), and the corresponding residual term (log Q_t — log Q_t*) with the increase of one variable at a time in the analysis. Observations for 1956 adjusted.

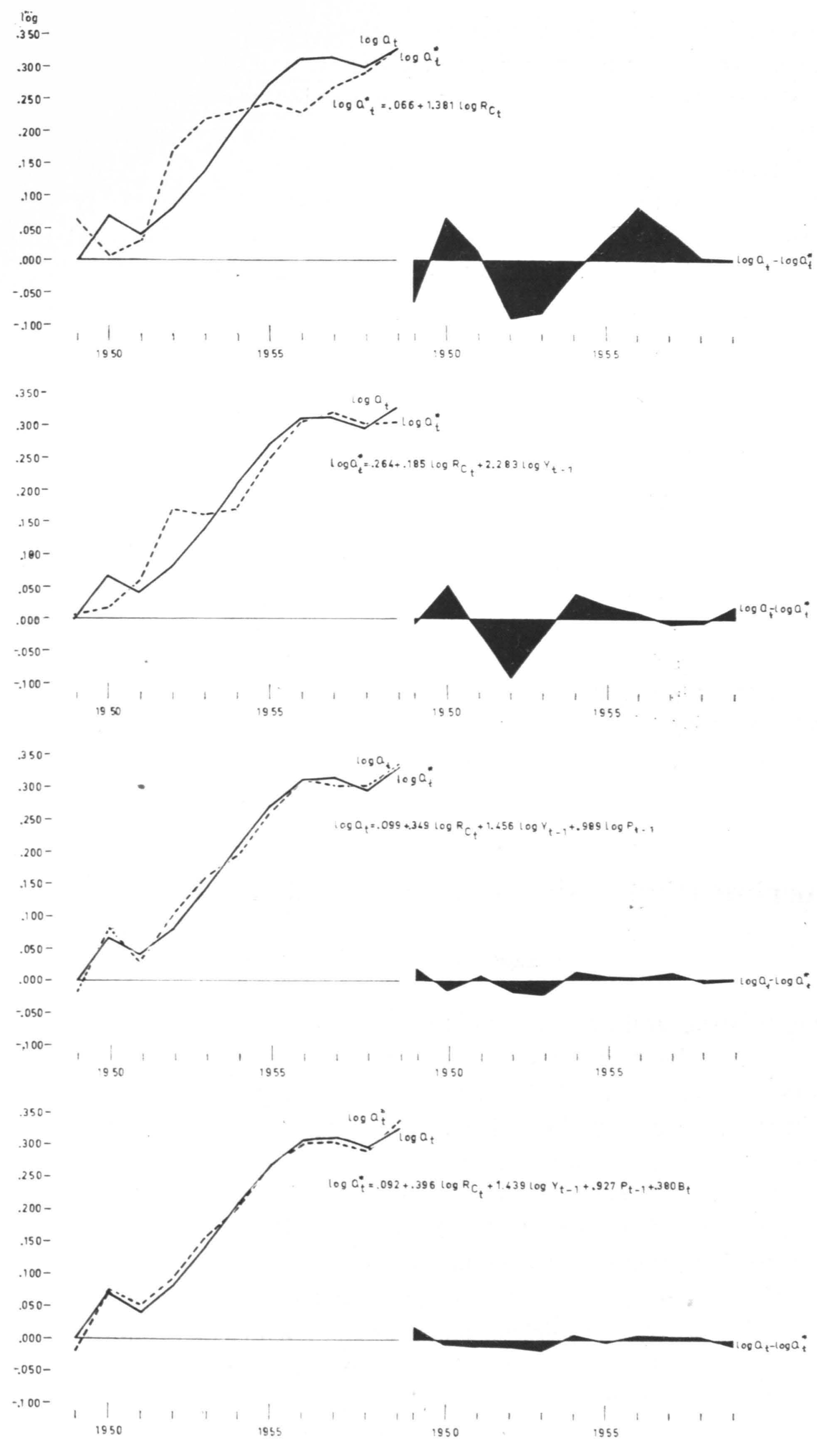


Figure 10. The logarithm of newsprint sales in Finland in 1949—59 (i) as observed (log Q_t) and (ii) as calculated from Model (26) (log Q_t*), and the corresponding residual term (log Q_t — log Q_t*) with the increase of one variable at a time in the analysis. Observations for 1956 adjusted.

The diminution of the significance of α_5 may be partly due to the fact that B_t was not adjusted for this analysis. It is difficult to determine the loss of exports due to the Strike with the aid of a seasonal index or similar means. Figures 9 and 10 show the development of the values of log Q_t* and the corresponding residual terms with the insertion of one variable at a time into the analysis, after correction of the observations for 1956, as indicated above. The »unexplained» variance of log Q_t in these models amounts to $\sim$ 1.2 per cent. The first autocorrelation coefficients of the residual terms for models (23) and (26), with observations for 1956 adjusted, were $r_{q_t q_{t-1}} = .0407 \pm .353$ and $r_{q_t q_{t-1}} = .117 \pm .349$, respectively.

The »adjusted» parameter estimates and coefficients of multiple correlation for models (23a) and (26a), based on the first logarithmic differences, were as follows:

Explanatory variable	Regression coefficient Model		
	(23a)	(26a)	α_i
	(with observations for 1956 adjusted)		
1	$-.001 \pm .010$	$-.003 \pm .009$	α_1
log $\Delta\ Y_{t-1}$	$1.560 \pm .443$	$1.383 \pm .361$	α_2
log $\Delta\ p_t$	$.341 \pm .198$	·	α_3
log $\Delta\ R_{C_t}$	·	$.448 \pm .194$	α_3
log $\Delta\ P_{t-1}$	$.897 \pm .181$	$.887 \pm .143$	α_4
log $\Delta\ B_t$	$.340 \pm .184$	$.388 \pm .165$	α_5
R	.877	.907	
s	.0170	.0149	

7 Experimenting with Models (23) and (26) Using Other Data

71 The Years 1927—1938

Owing to structural changes that take place in different observation periods, it is difficult to check the universality of models based on time series by testing their applicability to several observation periods. Nevertheless, such experiments are of interest since they throw light on the structural changes themselves and contrast the findings obtained here with those of other studies. In this study, models (23) and (26) were tried out with annual data for 1927—38. These data, however, were not fully comparable to those of 1949—59. The substitute series for subscription rates (p') was composed of the annual averages for four Helsinki papers (*Helsingin Sanomat, Hufvudstadsbladet, Sosialidemokraatti* and *Uusi Suomi*), weighted by the relative amounts of newsprint consumed by each in 1948 (see Koivulehto 1949, Appendices). Other weights were difficult to obtain. Owing to the different content of the subscription rates (p'), subscription revenue, R_{C_t}, has a somewhat different connotation here from that of the original model

(26). The business cycle variable was obtained by dividing the volume index of exports by its trend values as computed from the equation $X = 133.923 + 6.214\,t$. — Table VII (p. 78) shows the values of the variables used.[1]

Model

$$(23b)\quad \log Q_t = \alpha_1 + \alpha_2 \log Y_{t-1} + \alpha_3 \log p'_t + \alpha_4 \log P_{t-1} + \alpha_5 \log B_t + u_t,$$

was tried out first. On the basis of this trial, variables P_{t-1} and B_t were eliminated as insignificant and the following model was obtained:

$$(23b')\ \log Q_t = \alpha_1 + \alpha_2 \log Y_{t-1} + \alpha_3 \log p'_t + u_t.$$

In addition, the non-lagged price of newsprint (P_t) was tried out in the model, but proved to be nowhere near significant.

Certain conclusions can be drawn from the following tabulation, on comparing the parameter estimates of models (23b) and (23b') with those of model (23). It should be remembered that model (23), especially, contains a great deal of multicollinearity detrimental to the comparison. In the first place, it seems that the income elasticity of newsprint consumption (purchases) has diminished with the rising income level. This agrees well with the conclusion drawn by *FAO* (see p. 26, above). The above parameter estimates can also serve as a basis for computing the marginal propensity to consume newsprint at different levels of income. The latter indicates the ratio of consumption increase to the corresponding increase in income. Using the same symbols as in this study, the marginal propensity to consume newsprint (Q') can be expressed by the equation $Q' = \Delta Q : \Delta Y$.

Explanatory variable	Regression coefficient Model			
	(23)[2]	(23b)	(23b')	α_i
1	$.105 \pm .014$	$.042 \pm .013$	$.042 \pm .009$	α_1
log Y_{t-1}	$1.715 \pm .171$	$1.624 \pm .682$	$1.917 \pm .117$	α_2
log p_t	$.374 \pm .125$	·	·	α_3
log p'_t	·	$-.602 \pm .238$	$-.570 \pm .156$	α_3
log P_{t-1}	$.977 \pm .131$	$-.228 \pm .556$	·	α_4
log B_t	$.378 \pm .191$	$.092 \pm .173$	·	α_5
R	.994	.976	.981	
s	.0138	.0187	.0169	

Secondly, it should be remembered that the price of newsprint (P_{t-1}) had no significant coefficient in 1927—38, whereas in 1949—59 it had a positive one. One possible interpretation of this change is that, owing to the low income level in 1927—38, the publishers could not count on compensating for an increase in

[1] The trend subsequently computed for 1913—38 did not differ significantly from that for 1927—38. Official figures on the volume of exports before 1913 were not available.

[2] Observations for 1956 adjusted.

the price of newsprint by raising subscription and advertising rates (*cf.* p. 18, above). This conclusion is supported by the negative coefficient for subscription rates (p'_t) in 1927—38 as compared with the positive one for those (p_t) in 1949—59 — regardless of whether or not the values of these coefficients are influenced by multicollinearity. Thus an essential factor in structural changes seems to have been the ratio of the subscribers' income level to the subscription rates.

Model (26b) was set up by analogy with model (26):

$$\text{(26b)}\quad \log Q_t = \alpha_1 + \alpha_2 \log Y_{t-1} + \alpha_3 \log R'_{C_t} + \alpha_4 \log P_{t-1} + \alpha_5 \log B_t + u_t.$$

From the pertinent inversion stages of the moment matrix it may be inferred that the large standard errors of the parameter estimates for model (26b) are perhaps due to multicollinearity, especially that created by the variables P_{t-1} and B_t. These variables were therefore eliminated and the following model was obtained:

$$\text{(26b')}\quad \log Q_t = \alpha_1 + \alpha_2 \log Y_{t-1} + \alpha_3 \log R'_{C_t} + u_t.$$

The parameter estimates and other estimation statistics for models (26), (26b) and (26b') are tabulated below.

Explanatory variable	Regression coefficient Model (26)[1]	(26b)	(26b')	α_i
1	.092 ± .012	.042 ± .025	.047 ± .022	α_1
$\log Y_{t-1}$	1.439 ± .239	2.626 ± .776	2.406 ± .388	α_2
$\log R_{C_t}$	.396 ± .129	•	•	α_3
$\log R'_{C_t}$	•	−.034 ± .602	−.450 ± .314	α_3
$\log P_{t-1}$	.927 ± .126	.611 ± .798	•	α_4
$\log B_t$	.380 ± .188	.092 ± .271	•	α_5
R	.994	.954	.960	
s	.0136	.0258	.0239	

The results obtained from models (23b') and (26b') are largely parallel. From the inverse matrices, however, it can be concluded that model (26b') is considerably more multicollinear than model (23b'). This is due to the fact that, at the low income level of the time especially during the Great Depression in the early 1930's, newspaper subscription followed changes in income more closely than it does now that the income level is higher. It may also account in part for the differences in the results given by these two models.

[1] Observations for 1956 adjusted.

72 Forecasting with Model (26)

One of the prime requirements of a model is its ability to forecast. As an example, the capacity of model (26) to forecast newsprint sales in 1960 will be tried out below and compared with the actual sales of that year. The »adjusted» parameter estimates are given on page 51, above.[1] Table IV (p. 75) shows the observed values of the explanatory variables.

		$\alpha_1 = .09192$
$\log Y_{59} = .13033$	$\alpha_2 = 1.43870$	$\alpha_2 \log Y_{59} = .18751$
$\log P_{59} = .02531$	$\alpha_3 = .92670$	$\alpha_3 \log P_{59} = .02345$
$\log R_{C_{60}} = .18184$	$\alpha_4 = .39565$	$\alpha_4 \log R_{C_{60}} = .07194$
$\log B_{60} = .05308$	$\alpha_5 = .37997$	$\alpha_5 \log B_{60} = .02017$
		$\log Q_{60}^{\square} = .39499$

(= the logarithm of forecast sales per 15-or-more-year-old inhabitant in 1960)

Assuming that no structural changes occur between the period of observation and that of the forecast, and that the necessary observed values of the predetermined variables are known, the similarity between the forecast and the corresponding observation will depend on

1. the extent to which the parameter estimates approach their true values;
2. the extent to which the mean of the residual term approaches its expectation value of zero.

In these assumptions, the standard error of $\log Q_{60}^{\square}$ is obtained from the following equation (*cf.* Ezekiel 1941, p. 435)[1]:

$$\text{(27)}\quad \varepsilon^2_{\log Q_{60}^{\square}} = \frac{q_r^{(n)}}{f^{(n)}} \Big[1 + \frac{1}{f^{(1)}} + \alpha_{22} (\log R_{C_{60}})^2 + \alpha_{33} (\log Y_{59})^2 + \alpha_{44} (\log P_{59})^2 + \alpha_{55} (\log B_{60})^2 + 2\,\alpha_{23} \log R_{C_{60}} \log Y_{59} + 2\,\alpha_{24} \log R_{C_{60}} \log P_{59} + 2\,\alpha_{25} \log R_{C_{60}} \log B_{60} + 2\,\alpha_{34} \log Y_{59} \log P_{59} + 2\,\alpha_{35} \log Y_{59} \log B_{60} + 2\,\alpha_{45} \log P_{59} \log B_{60} \Big],$$

[1] The present calculation has been made to five decimal places.

[2] It can be proved that deletion of the first row and column (corresponding to the constant 1.000) of the inverse matrix (with n + 1 rows) obtained by Törnqvist's method gives the inverse of an n-rowed covariance matrix and provides the elements implied in Ezekiel's (1941, p. 345) formula (*cf.* Larna 1959 p. 182).

in which $q_r^{(n)}$ is the residual sum of the squares at the last inversion stage of the moment matrix (TÖRNQVIST's method), and $f^{(n)}$ is the number of degrees of freedom at the last inversion stage. The values of $\alpha_{22} \ldots \alpha_{55}$ have been obtained directly from the appropriate inversion stage of the moment matrix, while the observed values of the explanatory variables have been taken from Table IV (p. 75). The value of $\varepsilon_{\log Q_{60}^{\square}}$ here is $\sqrt{.011707} = .10820$.

$\therefore Q_{60}^{\square} = 2.48$, from which is obtained the forecast for newsprint sales per inhabitant in 1960: $9.386 \times 2.48 = 23.28$ kg. (see Table II, p. 73). The logarithmic standard error ($\varepsilon_{\log Q_{60}^{\square}}$) is $\sim \pm 1.28$ kg. Thus the difference between observed sales (21.87 kg. — see Table II, p. 73) and forecast sales is in the order of one standard error.

73 Extension of the Observation Period by one Year

When a model is used for forecasting it is desirable to increase the data underlying its parameter estimates according to the increase in the number of observations in the course of time. This is necessary in order to spot structural changes and or the emergence of new explanatory variables, if any. Furthermore, one observation must be added to the data at a time if it is intended to take full advantage of the observed values of the lagged variables, and thus reduce the extrapolating nature of forecast.

Changes in the parameter estimates of a model due to an additional observation can be calculated either by TÖRNQVIST's method using an electric desk calculator, in which case there is no need to recompute the moment matrix (see p. 32, above), or by an electronic computer. Since the time series in this case was quite short, we recomputed and inverted the moment matrix with an electronic computer at about the same cost as TÖRNQVIST's shorter manual method.

The observed values of C_t and p_t for 1960 were not directly available from statistics. C_{60} was formed from data supplied by the *Board of Posts and Telegraph* and *Helsingin Sanomat* (*cf.* p. 28, above). p_{60} was computed as a mean from data presented in Table IX (p. 79), weighted by the relative newsprint consumption of 17 newspapers in 1948. At this time, the papers were responsible for about 68 per cent of the total newsprint consumption in Finland (see KOIVULEHTO 1949, Appendices). To compute the index we utilized the corresponding mean for 1957, which had been computed in the same manner and the index figure for 1957 obtained from Table IV (p. 75). The observed values necessary for 1960 are set out in Table IV.

Worked out for the years 1949—60, the parameter estimates of models (23) and (26) (pp. 46 and 48, above) were as follows:

Explanatory variable	Regression coefficient Model (23) (1949—60 data)	(26)	α_i
1	.106 ± .016	.095 ± .013	α_1
$\log Y_{t-1}$	1.692 ± .203	1.395 ± .260	α_2
$\log p_t$	.322 ± .146	·	α_3
$\log R_{C_t}$	·	.388 ± .141	α_3
$\log P_{t-1}$	.962 ± .155	.924 ± .137	α_4
$\log B_t$	.188 ± .197	.240 ± .180	α_5
R	.992	.993	
s	.0165	.0149	

As can be seen, the business cycle variable has no significant coefficient for 1949—60. This may partly be accounted for by the fact that the observed value of the business cycle variable here is based on the trend for 1949—59. The changes in the other coefficients are perhaps normal for series of this length.

8 Summary and Conclusions

The market for newsprint produced in and sold to Finland is, practically speaking, a bilateral monopoly. It is fairly easy to list the most essential factors influencing the decisions of newspaper publishers in purchasing newsprint on the basis of information on the markets for newspapers and newsprint. The rhythm of decisions on prices and amounts purchased, and the classification of variables into predetermined and non-lagged endogenous, are also ascertainable. In its entirety the basic model is very complicated. Since data consist of annual items, the empirical applications are subject to several restrictions. To start with, the number of degrees of freedom is small, which calls for considerable simplification of the model. Simplification is desirable also because the data are multicollinear.

Two methods of selecting variables seemed to be available. According to the first, newsprint consumption is divided into components, each of which is considered separately (p. 24, above). A model on decision-making can be achieved with the aid of diagrams on the sequence of decision-making (Figures 1—3, pp. 21—24, above), though the price of newsprint is missing from it. Thus the model takes no account of the seller and his decisions. The second method, which proved more fruitful, starts out directly from factors influencing the publishers' decisions in purchasing newsprint, eliminating the least significant intuitively and simultaneously determining the lags. From the remaining variables, simple models were specified along two different lines — based either on the economy of newspapers (Chapter 628) or on the economy of the newspapers' customers (Chapter 629.2). In addition, a few models were derived by replacing certain variables by others which explained them (Chapter 629.1).

A logarithmic function was used to describe the statistical relationship between the variable to be explained and each explanatory variable. The most promising results were obtained from the models based on the economy of the newspapers' customers [models (23) and (26)]. It was also attempted to reduce the collinearity of the variables by calculating the parameter estimates by means of the first logarithmic differences of the observations [models (23a) and (26a)]. This did not eliminate multicollinearity, but the analyses performed tended to show that the multicollinearity in models (23) and (26), for example, was not due to a linear trend common to all the variables.

Sales of newsprint in the year t can be forecast by means of models (23) and

(26) provided that two observed values for the year t are available as explanatory variables; the other two explanatory variables are in the models with one-year lags. As regards the non-lagged variables in models (23) and (26), the average subscription rate for the newspapers is known towards the end of the year $t-1$, while C_t, the component variable of R_{C_t}, can be estimated with reasonable accuracy on the basis of advance subscriptions paid by the end of the year $t-1$. The business cycle variable B_t, must be forecast; data on it will probably improve with continued research.

In view of the possibility of structural changes, it is desirable to reestimate the parameters as the number of observations increases. This can be done either by a manual procedure according to TÖRNQVIST's method (see p. 32, above), in which case there is no need to recompute the moment matrix, or by means of an electronic computer using a standard programme.

The models developed in this study are capable of forecasting potential consumption. The economically logical basis implicit in the classification of forecasts given in the introduction may sometimes be of secondary importance when set against statistical criteria. For example, a forecast of actual consumption may have a considerably larger variance than a forecast of potential consumption merely because it is seldom possible to identify demand and supply curves (except formally) and receive a high total explanation. On the other hand, where identification succeeds to the point of producing structural parameters, most forecasts based on such models are free of errors due to subsequent changes in the intercorrelations between explanatory variables. Even the best models in this study are multicollinear. The further into the future a forecast is extended, the greater is the possibility that relationships between the explanatory variables will change. In a forecast one year ahead this danger is almost insignificant provided the parameters of the model are kept up to date with an increasing number of observations. On the other hand, even structure-identifying models are not free from the structural changes that can emerge in the course of time. The basic model may be quite universal, but the parameter estimates obtained from its application are valid only for a given time and place.

In principle it is possible — where data permit — to construct models that provide paramenters of action for the economic units concerned. This is not the aim here. Instead, it is intended that the newsprint seller will profit from the models achieved by using them to forecast sales in a future period. In this way, he can avoid both loss of interest due to acquiring a surplus of raw material or acute shortages of raw material due to underestimating short term requirements — a mistake that can cost the producer dearly owing to the consequent emergency measures necessary and, sometimes, lost sales opportunities. In Finland the latter factor is less important, because the domestic market for newsprint forms only a minor part of the total; the bulk goes for export. Elsewhere, this may not be the case.

It is conceivable that wood utilization as a whole can be studied on the basis of factors affecting each end use. The object in this study has been forecasting, but it could just as well have been causal explanation, with the emphasis on future decisions in economic policy.

Finally, the logicality of the hypotheses expressed in Figures 1—3 (pp. 21—24, above) may be considered in the light of the results obtained. We could, of course, argue that their logicality cannot be questioned, since the hypotheses were based on pre-existing economic theories. The failure or success of an economic theory to describe a given phenomenon does not invalidate the theory being considered. No economic laws are universal — *i.e.* independent of time and place. From the standpoint of scientific methodology, they are procedural rules — schemes of interpretation. It is therefore more rational to ask whether we have selected the best theory as the basis for empirical analysis. Figures 1—3 represent a complex of theories and it is obvious from the results obtained that some of the hypotheses based on these theories are better than others for the present study. As stated earlier, those based on the economy of the newspapers' customers provide the best fit in empirical analysis.

Literature Cited

ABBOTT, LAWRENCE. 1955. Quality and Competition. An Essay in Economic Theory. New York, N.Y.

AITCHISON, T. and BROWN, C. 1957. The Log-Normal Distribution. Cambridge.

ANDERSON, ARTHUR G., MANDEVILLE, MORTEN JOSEPH, and ANDERSON, JOHN MUELLER. 1942. Industrial Management. New York, N.Y.

BARTIZAL, JOHN R. 1940. Budget Principles and Procedure. New York, N.Y.

BENTZEL, RAGNAR. 1953. Inkomstfördelningen i Sverige. Summary: The Distribution of Income in Sweden. Uppsala.

BERG, E. A. 1952. Sanomalehtitoimen tuotannontekijäin riippuvaisuussuhteista. Helsinki.

BETHEL, LAWRENCE L., ATWATER, FRANKLIN S., SMITH, GEORGE H. E., and STACKMAN, HARVEY A. 1956. Industrial Organization and Management. New York, N.Y. — Toronto — London.

BOULDING, KENNETH E. 1955. Economic Analysis (*3rd* ed.). London.

CHERNOFF, H. and RUBIN, H. 1953. Asymptotic Properties of Limited-Information Estimates under Generalized Conditions. Studies in Econometric Method by Cowles Commission Research Staff Members. Edited by Wm. C. Hood and Tjalling C. Koopmans. Cowles Commission for Research in Economics Monograph No. 14, pp. 200—212. New York, N.Y. — London.

DUERR, WILLIAM and VAUX, HENRY J. (ed.). 1953. Research in the Economics of Forestry. Washington, D.C.

ERVASTI, SEPPO. 1958. Suomen metsäteollisuuden laajentamismahdollisuudet tuotteiden käytön kehityksen kannalta. Summary: The Possibilities of Expanding Finland's Forest Industry from the Standpoint of Developing Product Utilization. Reprint from Silva Fennica 97. Helsinki.

European Timber Trends and Prospects. 1953. A Study Prepared Jointly by the Secretariats of the Food and Agriculture Organization of the United Nations and the United Nations Economic Commission for Europe. Geneva.

EZEKIEL, MORDECAI. 1941. Methods of Correlation Analysis (*2nd* ed.). New York, N.Y. — London.

FOOTE, RICHARD J. 1958. Analytical Tools for Studying Demand and Price Structures. U.S. Department of Agriculture. Agriculture Handbook No. 146. Washington, D.C.

FOOTE, RICHARD J. and FOX, KARL A. 1954. Analytical Tools for Measuring Demand. United States Department of Agriculture. Agriculture Handbook No. 64. Washington, D.C.

FRISCH, RAGNAR. 1934. Statistical Confluence Analysis by Means of Complete Regression Systems. Publication No. 5 of the University Institute of Economics. Oslo.

GREGORY, G. ROBINSON. 1960. A Statistical Investigation of Factors Affecting the Market for Hardwood Flooring. Forest Science No. 2/1960, pp. 123—134. Washington, D.C.

HAIKALA, EINO. 1956. Maatalouden ominaissuhdanteet ja cobwebteoria. Summary: On the Specific Cycles of Agriculture and the Cobweb Theorem. Helsinki.

HICKS, J. R. 1932. Theory of Wages. London.

HOLOPAINEN, VILJO. 1960. On the Price Elasticity of the Supply of Sawnwood for Export. Summary in Finnish: Sahatavaran vientitarjonnan hintajousto. Reprint from Acta forestalia Fennica 73. Helsinki.
HURWICZ, LEONID. 1950. Prediction and Least Squares. Statistical Inference in Dynamic Economic Models by Cowles Commission Research Staff Members and Guests. Edited by Tjalling C. Koopmans. Cowles Commission for Research in Economics Monograph No. 10, pp. 266—300. New York, N.Y.—London.
KLEIN, LAWRENCE R. 1953. A Textbook of Econometrics. Evanston, Ill. — White Plains, N.Y.
—»— 1960. Single Equation *vs.* Equation System Methods of Estimation in Econometrics. Econometrica No. 4/1960, pp. 866—871. Menasha, Wis.
KOIVULEHTO, A. 1949. Kertomus paperin säännöstelystä. [Unpublished typescript]. Helsinki.
LARNA, KAARLO. 1959. The Money Supply, Money Flows and Domestic Product in Finland, 1910—1956. Economic Studies XXIII. Helsinki.
LAURILA, EINO H. 1956. Yleislakon aiheuttamista taloudellisista menetyksistä. Kansallis-Osake-Pankki, Taloudellinen katsaus No. 1/1956, pp. 38—39.
Lehtien postimaksukomitean mietintö. 1959. Helsinki.
MATTINEN, PIRKKO. 1955. Kilpailun tulevaisuus I. Liiketaloudellinen aikakauskirja (The Journal of Business Economics) No. 1/1955, pp. 16—33. Helsinki.
MOOD, ALEXANDER MCFARLANE. 1950. Introduction to the Theory of Statistics. New York, N.Y.
NICHOLLS, WILLIAM H. 1941. A Theoretical Analysis of Imperfect Competition with Special Application to the Agricultural Industries. Ames, Ia.
NIITAMO, OLAVI. 1958. Tuottavuuden kehitys Suomen teollisuudessa vuosina 1925—1952. Summary: The Development of Productivity in Finnish Industry, 1925—1952. Kansantaloudellisia tutkimuksia (Economic Studies) XX. Helsinki.
NIITAMO, OLAVI and PULLIAINEN, KYÖSTI. 1960. Taloudellinen malli. Summary: Economic Model. Reprint from Kansantaloudellinen aikakauskirja (The Finnish Economic Journal) No. 4/1960. Helsinki.
The Outlook for the Canadian Forest Industries. 1957. By Davis, John, Best, A.L., Lachance, P. E., Pringle, S. L., Smith, J. M., and Wilson, D. A. Royal Commission on Canada's Economic Prospects. [Hull].
Paper for Printing Today and Tomorrow. 1952. By the Intelligence Unit of The Economist, London· Presented jointly by UNESCO and FAO. Paris.
Paper for Printing and Writing. 1954. Tentative Forecasts of Demand in 1955, 1960 and 1965. By the Intelligence Unit of The Economist, London. Clearing House Department of Mass Communication. Reports and Papers on Mass Communication No. 12. London.
Possibilities for the Development of the Pulp and Paper Industry in Latin America. 1954. A Joint Study by the Economic Commission for Latin America and the Food and Agriculture Organization of the United Nations. New York, N.Y.
Posti- ja lennätinlaitos. [Different years]. Kotimaisten sanomalehtien hinnasto. Helsinki.
Potential Requirements for Timber Products in the United States. 1946. Report II from a Reappraisal of the Forestry Situation. U.S. Forest Service. Washington, D.C.
Pulp and Paper Prospects in Latin America. 1955. First part: Report for the Latin American Meeting of Experts on the Pulp and Paper Industry Sponsored by the Secretariats of the Economic Commission for Latin America, the Food and Agriculture Organization of the United Nations, and the Technical Assistance Administration. Second part: Working Paper Submitted to the Meeting. New York, N.Y.
Pulp, Paper and Board Supply-Demand. 1957. Report of the Committee on Interstate and Foreign Commerce. Union Calendar No. 198. House Report No. 573. Washington, D.C.
PÖYHÖNEN, PENTTI. 1955. Ekonometrinen tutkimus tonttien hinnoista. Summary: An Econometric Investigation of City Land Market Prices. Valtion teknillinen tutkimuslaitos (The State Institute for Technical Research, Finland). Julkaisu (Publication) No. 31. Helsinki.

QUIST, MARTTI. 1960. Lukemisharrastus. Helsingissä suoritettu haastattelututkimus. [Hämeenlinna].
RIIHINEN, PÄIVIÖ. 1957. Etelä-Amerikan havusahatavarataseen ennuste. Markkinatieteellinen tutkimus tase-erotuksesta. [Unpublished licentiate paper]. Helsinki.
—»— 1958. Tuotannon tavoite metsätaloudessa. Teoreettisten mallien tarkastelu. Summary: Production Goal in Forestry: an Evaluation of Theoretical Models. Reprint from Kansantaloudellinen aikakauskirja (The Finnish Economic Journal) No. 2/1958. Helsinki.
SARIOLA, SAKARI. 1954. Lappi ja väkijuomat. Helsinki.
SHAMES, LEE M. 1946. A Forecast of Lumber Demand. Journal of Forestry No. 7/1946, pp. 477—487. Washington, D.C.
STIGLER, GEORGE J. 1952. The Theory of Price (rev. ed.). New York, N.Y.
STONE, RICHARD and ROWE, D. A. 1955. Personal and Corporate Spending and Saving Functions with Application to the United Kingdom. [Report at] *4th* Conference of the International Association for Research in Income and Wealth, Hindsgavl, Denmark, 6—13 September, 1955.
Suhdannesarjat. [Different years]. Valtiovarainministeriön kansantalousosasto (Ministry of Finance, Economic Department). Helsinki.
SVT[1] [Different years]. Teollisuustilastoa (Industrial Statistics of Finland). Helsinki.
—»— [Different years]. Tilastollinen vuosikirja (Statistical Yearbook of Finland). Helsinki.
—»— [Different years]. Ulkomaankauppa (Foreign Trade). Helsinki.
THEIL, H. 1954. Estimation of the Parameters of Econometric Models. International Statistical Institute. Bulletin No. 34, pp. 122—128.
—»— 1958. Economic Forecasts and Policy. Assisted by Cramer, J. S., Moerman, H. and Russchen, A. Contributions to Economic Analysis XV. Amsterdam.
Tilastokatsauksia (Bulletin of Statistics). [Different years]. Tilastollinen päätoimisto (Central Statistical Office in Finland). Helsinki.
Timber Resources for America's Future. 1958. U.S. Department of Agriculture, Forest Service. Forest Resource Report No. 14. Washington, D.C.
TINTNER, GERHARD. 1952. Econometrics. New York, N. Y. — London.
—»— 1960. Handbuch der Ökonometrie. Enzyklopädie der Rechts- und Staatswissenschaft. Berlin —Göttingen—Heidelberg.
TÖRNQVIST, Leo. 1957. A Method for Calculating Changes in Regression Coefficients and Inverse Matrices Corresponding to Changes in the Set of Available Data. Skandinavisk Aktuarietidskrift No. 3—4/1957, pp. 219—226.
—»— 1958. Utvecklingen av industriproduktionen i Finland åren 1925—56. Ekonomiska samfundets tidskrift No. 3/1958, pp. 150—159. Helsingfors.
VALAVANIS, STEFAN. 1959. Econometrics. An Introduction to Maximum Likelihood Methods. New York, N.Y. — Toronto—London.
VAUX, HENRY J. 1950. An Economic-Statistical Analysis of Lumber Requirements for California Housing. Hilgardia No. 16/1950. Berkeley, Calif.
—»— 1954. Economics of the Young-Growth Sugar Pine Resource. California Agricultural Experiment Station. Bulletin No. 738. Berkeley, Calif.
VAUX, HENRY and ZIVNUSKA, JOHN A. 1952. Forest Production Goals: A Critical Analysis. Reprinted from Land Economics No. 4/1952. Madison, Wis.

[1] SVT = Suomen virallinen tilasto *(Officical Statistics of Finland)*

WOLD, HERMAN. 1953. (in association with Lars Juréen.) Demand Analysis. A Study in Econometrics. New York, N.Y. — Stockholm.
—»— 1956. Causal Inference from Observational Data: A Review of Ends and Means. Journal of the Royal Statistical Society No. 1/1956, pp. 28—50. London.
World Consultation on Pulp and Paper Demand, Supply and Trade, Rome, 4—19 September, 1959. A Forecast of Demand — the Period to 1975. Food and Agriculture Organization of the United Nations. [Mimeographed]. Rome.
World Demand for Paper to 1975. 1960. Food and Agriculture Organization of the United Nations. Rome.
World Pulp and Paper Resources and Prospects. 1954. A Survey Prepared by the Food and Agriculture Organization in Co-operation with the Secretariats of the United Nations Educational, Scientific and Cultural Organization (UNESCO); the Economic Commission for Europe (ECE); and the Economic Commission for Latin America (ECLA). New York, N.Y.
ZEUTHEN, F. 1930. Problems of Monopoly and Economic Warfare. London.
—»— 1955. Economic Theory and Method. Cambridge, Mass.

SELOSTUS:

SANOMALEHTIPAPERIN MYYNTI SUOMESSA VUOSINA 1949—1959

TUTKIMUS LYHYTJÄNTEISTEN ENNUSTEIDEN MALLEISTA

Tässä tutkimuksessa pyritään kehittämään Suomessa tuotetun ja Suomeen myydyn sanomalehtipaperin määrän vuosittaisia vaihteluja kuvaava malli, jossa matemaattinen yksinkertaisuus, kuvauksen tarkkuus ja yleistettävyys mahdollisuuksien mukaan yhdistyvät.

Suomessa tuotetun sanomalehtipaperin markkinoilla kotimaassa vallitsee bilateraalista monopolia lähenevä kilpailun muoto. Keskeisimmät sanomalehtien paperinostopäätöksiin vaikuttavat tekijät ovat verrattain helposti lueteltavissa sanomalehtien ja paperin markkinoista saatujen tietojen perusteella. Myös hintaa ja ostettavaa paperimäärää koskevien päätösten rytmi sekä muuttujien jako ennalta määrättyihin ja viivästymättömiin endogenisiin ovat selvitettävissä. Perusmalli kokonaisuudessaan on melko monimutkainen. Aineiston muodostuessa vuosittaisista tiedoista on käytännöllisillä sovellutuksilla useita rajoituksia. Vapausasteita on alkuun lähdettäessä vähän, mikä pakottaa mallin voimakkaaseen yksinkertaistamiseen. Sen yksinkertaistamiseen on syytä myös siksi, että aineisto on multikollineaarista.

Muuttujien valintaan sovellutuksia varten näytti olevan tarjolla kaksi menetelmää. Toisessa näistä jaetaan sanomalehtipaperin käyttö komponentteihin ja tarkastellaan näistä kutakin erikseen (ks. s. 24 ed.). Käyttäen apuna päätöksentekojärjestystä esittäviä kaavioita (kuvat 1—3, s. 21—24) päästään päätöksentekijämalliin, josta kuitenkin puuttuu sanomalehtipaperin hinta. Täten saatu malli ei siis ota huomioon sanomalehtipaperin myyjän hintapäätöksiä. Toinen menetelmä, joka osoittautui tehokkaammaksi, lähtee erittelemään suoraan sanomalehtien paperinostopäätöksiin vaikuttavia tekijöitä karsien näistä intuitiivisesti vähämerkityksiset ja määrittäen viivästykset samanaikaisesti. Näin jäljelle jääneistä muuttujista spesifioitiin yksinkertaisia malleja kahdella eri perusteella: lähtemällä vaihtoehtoisesti sanomalehtien (kappale 628) tai näiden palvelusten tarvitsijain taloudesta (kappale 629.2). Sanomalehtien taloudesta

[1] Tutkimuksessa käytettyjen symbolien luettelo on esitetty liitteessä (s. 70—71).

johdettiin lisäksi muutamia malleja korvaamalla tiettyjä muuttujia niitä selittävillä toisilla muuttujilla (kappale 629.1).

Selvitettävän ja kunkin selvittävän muuttujan välistä tilastollista yhteyttä kuvattiin logaritmisella funktiolla. Lupaavimmat tulokset saatiin sanomalehtien tarvitsijain talouteen perustuvista malleista [mallit (23), (23b), (26) ja (26b)]. Muuttujien kollineaarisuutta pyrittiin vähentämään laskemalla parametrien estimaatit myös havaintojen logaritmien ensimmäisillä erotuksilla. Multikollineaarisuus ei mallista täten poistunut, mutta suoritetut analyysit pyrkivät osoittamaan, että multikollineaarisuus ei esimerkiksi malleissa (23) ja (26) johdu muuttujien lähes yhtäläisestä suoraviivaisesta trendistä.

Malleilla (23) ja (26) voidaan ennustaa sanomalehtipaperin myynti vuonna *t* tarvitsematta tuntea muiden kuin kahden muuttujan havaintoarvot vuonna *t*. Muut kaksi selvittävää muuttujaa ovat malleissa yhden vuoden viivästyksin. Mallin (23) viivästymättömistä muuttujista sanomalehtien keskimääräinen tilaushinta tunnetaan vuoden *t-1* lopulla, kun taas mallissa (26) esiintyvä vuoden *t* levikki (C_t) voitaneen melko tarkasti arvioida vuoden *t-1* loppuun mennessä tehtyjen ennakkotilausten perusteella. Suhdannemuuttuja (B_t) on ennustettava; siitä saatavat ennakkotiedot paranevat suhdannetutkimuksen edistyessä.

Rakennemuutosten silmällä pitämiseksi on mallin parametrit syytä estimoida uudelleen sitä mukaa kuin havaintoja karttuu. Tämä käy päinsä joko käsin laskien Törnqvistin menetelmällä (ks. s. 32.) tarvitsematta laskea momenttimatriisia uudelleen tai vakio-ohjelmaa käyttäen matematiikkakoneella.

Tässä tutkimuksessa kehitetyillä malleilla voidaan ennustaa sanomalehtipaperin potentiaalinen käyttö. Johdannossa esitetty taloustieteellisesti looginen ennusteiden luokittelu saattaa joskus olla toisarvoinen tilastollisten kriteerien rinnalla. Niinpä markkinakäytön ennusteella saattaa olla huomattavasti suurempi varianssi kuin potentiaalisen käytön ennusteella pelkästään siksi, että on harvoin mahdollista identifioida (paitsi muodollisesti) kysyntä- ja tarjontakäyrät ja samalla saavuttaa korkea kokonaisselitys. Toisaalta, missä identifionti onnistuu niin hyvin, että rakenneparametrit voidaan määrittää, ovat tällaisiin malleihin perustuvat ennusteet todennäköisesti vapaita selvittävien muuttujien keskinäisissä korrelaatioissa myöhemmin sattuvista muutoksista. Parhaatkin mallit tässä tutkimuksessa ovat multikollineaarisia. Mitä pitemmälle tulevaisuuteen ennuste ulottuu, sitä todennäköisempää on, että selvittävien muuttujien väliset tilastolliset yhteydet muuttuvat. Ennustettaessa vuosi kerrallaan tämä vaara on miltei vailla merkitystä, mikäli mallin parametrien estimaatit pidetään havaintojen lisääntyessä ajan tasalla. Toisaalta eivät rakenteita identifioivat mallitkaan ole vapaita ajan mukana sattuvista rakennemuutoksista. Perusmalli saattaa olla varsin pitkälle yleistettävä, mutta sen sovellutuksista saadut parametrien estimaatit pätevät ainostaan tiettynä aikana määrätyllä alueella.

Periaatteessa on mahdollista — missä aineisto sallii — rakentaa malleja, joista saadaan tarkasteltavien talousyksiköiden (keskeisimmät) toimintaparametrit. Se ei kuitenkaan ole tämän tutkimuksen tarkoitus. Sen sijaan tässä tutkimuksessa kehitettyjä malleja voi käyttää hyväkseen sanomalehtipaperin myyjä ennustaessaan tulevan myynnin. Tällä tavoin se voi välttää liian suurista raaka-aineen hankinnoista johtuvan korkotappion tai läheisen tulevaisuuden tarpeiden aliarvioinnista aiheutuvan raaka-aineen äkillisen puutteen; tämä erehdys saattaa aiheuttaa tuottajalle suunnitelmattomuudesta ja käyttämättä jääneistä myyntimahdollisuuksista johtuvia menetyksiä. Suomessa viimeksi mainittu näkökohta on vähemmän tärkeä, koska kotimaiset markkinat muodostavat pienen osan täällä tuotetun sanomalehtipaperin markkinoista. Muualla ei ehkä ole näin.

On helppo todeta, että puun käyttöä kokonaisuudessaan voidaan tutkia kuhunkin käyttömuotoon vaikuttavien tekijöiden perusteella. Tämän tutkimuksen tarkoitus on ennustaminen, mutta se olisi yhtä hyvin voinut olla kausaalinen selittäminen — erityisesti tulevassa talouspolitiikassa tehtäviä päätöksiä silmällä pitäen.

Lopuksi voidaan tarkastella kuvissa 1—3 (s. 21—24) esitettyjen hypoteesien paikkansa pitävyyttä saavutettujen tulosten valossa. Voitaisiin tietenkin väittää, että niiden mielekkyyttä ei voida asettaa kyseenalaiseksi, koska ne perustuvat aiemmin voimassa oleviin talousteorioihin. Talousteorian pätevyys tai pätemättömyys kuvattaessa tiettyä ilmiötä ei anna tälle teorialle yleispätevyyttä eikä myöskään kumoa sitä. Taloudelliset lainmukaisuudet eivät ole universaaleja, riippumattomia ajasta ja paikasta. Tieteen metodologian kannalta ne ovat menettelysääntöjä — tulkintakaavoja, jotka osoittavat, miten todellisuutta on kuvattava. On sen vuoksi mielekkäämpää kysyä, onko valittu paras käytettävissä oleva teoria empiirisen analyysin pohjaksi. Kuvat 1—3 edustavat teoriakompleksia, ja on ilmeistä, että muutamat näihin teorioihin perustuvista hypoteeseista ovat tämän tutkimuksen kannalta »parempia» kuin toiset. Parhaiten empiiristä aineistoa »selittävät» sanomalehtien asiakkaiden talouteen perustuvat mallit.

Symbols Used in the Study[1]

Tutkimuksessa käytetyt symbolit

Q_t = quantity of newsprint sold per 15-or-more-year-old inhabitant (in the year t) — *myydyn sanomalehtipaperin määrä 15 vuotta täyttänyttä asukasta kohden (vuonna t)*

Y_{t-1} = real income of subscribers, advertisers and other customers per 15-or-more-year-old inhabitant — *tilaajien, ilmoittajien ja muiden asiakkaiden reaalitulo 15 vuotta täyttänyttä asukasta kohden*

A_{t-1} = consumption of advertising space per 15-or-more-year-old inhabitant — *ilmoitustilan käyttö 15 vuotta täyttänyttä asukasta kohden*

π_{t-1} = average unit price of advertising space — *ilmoitustilan keskimääräinen yksikköhinta*

P_{t-1} = price of newsprint — *sanomalehtipaperin yksikköhinta*

p_t = average (annual) subscription rate — *vuositilauksen keskihinta*

p'_t = average (annual) subscription rate for four Helsinki papers — *neljän Helsingin lehden vuositilauksen keskihinta*

C_t = circulation per 15-or-more-year-old inhabitant — *levikki 15 vuotta täyttänyttä asukasta kohden*

$R_{A_{t-1}}$ = revenue from advertisements per 15-or-more-year-old inhabitant — *ilmoitustulot 15 vuotta täyttänyttä asukasta kohden*

R_{C_t} = revenue from subscriptions per 15-or-more-year-old inhabitant — *lehtimyyntitulot 15 vuotta täyttänyttä asukasta kohden*

V = average size of one copy — *numerokappaleen keskikoko*

H = population — *väestö*

N = space occupied by news and features — *uutistilan käyttö*

t = time — *aika*

B_t = business cycle variable — *suhdannemuuttuja*

E = economic activity of advertisers — *ilmoittajien taloudellinen aktiviteetti*

E' = economic activity of the service industries — *palveluselinkeinojen taloudellinen aktiviteetti*

E'' = economic activity of customers and clients of the service industries — *palvelusten tarvitsijain taloudellinen aktiviteetti*

T = volume of trade, industry and house building — *kaupan, teollisuuden ja talonrakennustoiminnan volyymi*

L = unemployment — *työttömyys*

K = level of knowledge — *tiedon taso*

S = volume of retail and gross sales — *vähittäis- ja tukkumyynnin volyymi*

Z = consumption of news and advertising space — *uutis- ja ilmoitustilan käyttö*

G_{t-1} = level of cost of the newspapers — *sanomalehtien kustannustaso*

u_t = random variable — *satunnaismuuttuja*

[1] For units of measurement and other technical details, see pp. 27—30, above.

$q_r(^1)$ = residual sum of the squares at the first inversion stage of the moment matrix — *jäännösneliösumma momenttimatriisin ensimmäisessä kääntövaiheessa*

q_t = residual term ($\log Q_t - \log Q_t^*$) — *jäännöstermi* ($\log Q_t - \log Q_t^*$)

$f(^1)$ = number of degrees of freedom at the first inversion stage of the moment matrix — *vapausasteiden luku momenttimatriisin ensimmäisessä kääntövaiheessa*

a_i = estimate of the parameter obtained as a statistical constant or as a regression coefficient of a given variable — *tilastollisena vakiona tai määrätyn muuttujan regressiokertoimena saatu parametrin estimaatti*

a_{ij} = element of the inverse matrix, *ith* row, *jth* column — *käänteismatriisin alkio, i:s rivi, j:s sarake*

ε_{a_i} = standard error of a_i — *a_i:n keskivirhe*

s = standard error of estimate (from a regression equation) — *(regressioyhtälöstä saadun) estimaatin keskivirhe*

R = coefficient of multiple correlation — *yhteiskorrelaatiokerroin*

Q' = marginal propensity to consume newsprint — *sanomalehtipaperin käytön raja-alttius*

Q_t^* = newsprint sales per 15-or-more-year-old-inhabitant as calculated from a model — *mallilla laskettu sanomalehtipaperin myynti 15 vuotta täyttänyttä asukasta kohden*

$Q_t^\square$ = newsprint sales per 15-or-more-year-old inhabitant as forecast by means of a model — *mallilla ennustettu sanomalehtipaperin myynti 15 vuotta täyttänyttä asukasta kohden*

Appendix Tables

Table I. Total output, sales of newsprint on domestic market, 1949—1960.[1]

Calendar year	Output	Sales to Finland	Sales to Finland /Output
	1,000 tons		Per cent
1949	380	26	6.8
1950	403	31	7.7
1951	410	29	7.1
1952	431	32	7.4
1953	438	37	8.4
1954	445	44	10.0
1955	526	51	9.6
1956	597	53	8.8
1957	626	58	9.2
*1958	613	56	9.2
*1959	633	61	9.7
*1960	753	67	8.9

Sources: — SVT., *Teollisuustilastoa.*
— *Tilastokatsauksia* No. 1/1961, p. 6.
— Table II.

[1] The percentages in the table have been calculated from unrounded figures.

Table II. Basic values of variables given in Table IV.

Calendar year	Sales of newsprint	Population, 15 years of age or over	Sales of newsprint per 15-or-more-year-old inhabitant	Price of newsprint	Index of real national product	Real national product per inhabitant [1]	Consumption of advertising space	Consumption of advertising space per inhabitant	Number of copies distributed by mail	Number of copies distributed by mail per inhabitant	Annual means of subscription rates	Annual means of advertising space rates	Price index of net national product	Price of newsprint deflated by the NNP[2] price index	Volume index of retail and gross sales	Volume index of exports
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	1,000 kg.	1,000	kg.	Fmk. /kg.	1938 = 100	1949 = 100	1949 = 100	1949 = 100	Millions of copies	1949 = 100	1938 = 100	1938 = 100	1926 = 100	Fmk. /kg.	1949 = 100	1954 = 100
1948	27,235	2,793	9.75	11.87	111	98	107	107	438	104	445	1,061	963	12.33	92	70
1949	26,328	2,805	9.39	15.29	114	100	100	100	422	100	690	1,073	967	15.81	100	65
1950	30,820	2,816	10.94	15.29	119	104	113	113	452	107	692	1,253	1,171	13.06	115	69
1951	29,105	2,828	10.29	19.58	131	114	119	118	464	109	968	1,589	1,581	12.38	125	87
1952	32,327	2,849	11.35	22.42	130	112	136	134	488	114	1,187	1,658	1,595	14.06	137	77
1953	37,075	2,878	12.88	23.70	132	113	144	140	514	119	1,210	1,668	1,573	15.07	130	86
1954	44,207	2,911	15.19	24.84	144	122	173	167	536	122	1,212	1,668	1,592	15.60	146	100
1955	51,454	2,941	17.49	27.14	154	129	202	193	547	124	1,297	1,863	1,691	16.05	176	109
1956	53,117	2,977	17.84	28.61	156	129	196	185	542	121	1,336	2,163	1,880	15.22	181	107
1957	58,209	3,009	19.34	29.86	157	128	199	185	577	128	1,504	2,320	1,941	15.38	170	117
1958	55,877	3,008	18.58	33.37	157	128	204	190	565	125	1,720	2,615	2,067	16.14	164	115
1959	60,857	3,039	20.02	34.62	167	135	220	203	553	121	1,886	2,675	2,057	16.83	184	131
1960	66,715	3,050	21.87	35.38	181	146	248	228	579	126	1,858	. .	2,150	16.46	. .	151

Sources: Column 1, 4 — *Paperikonttori's* price adjustment sheets.
2 — *Tilastokatsauksia* No. 12/1959, p. 42.
3 — *Tilastokatsauksia.*
7 — *Suhdannesarjat.*
9 — *Lehtien postimaksukomitean mietintö* 1959, p. 14 (See text, p. 28, above).
10—12 — *Lehtien postimaksukomitean mietintö* 1959, p. 69.
13 — *Bank of Finland Institute for Economic Research.*
15 — *Central Statistical Office.*
16 — *Tilastokatsauksia.*

[1] See footnote, p. 37.
[2] Net national product.

Table III. Computation of index of the cost level (G) of newspapers, 1949—1959. (Ref. p. 29, above).

Calendar year	Total salaries in the printing industries	Column 1 deflated by the NNP[1] price index	Total salaries in the printing industries	Total wages of wage earners working for newspapers	Column 4 deflated by the NNP price index	Cost of newsprint in newspapers deflated by the NNP price index	Columns 5 and 6 together	Column 7	.27 × column 3 + .57 × column 8	Cost of production per kilo of newsprint purchased (G)
	1	2	3	4	5	6	7	8	9	10
	Millions of Fmk.	Millions of Fmk.	1949 = 100	Millions of Fmk.	Millions of Fmk.	Millions of Fmk.	Millions of Fmk.	1949 = 100		1949 = 100
1948	297.7	309.1	87.4	91.4	94.9	335.8	430.7	82.9	70.851	82
1949	342.1	353.8	100.0	99.8	103.2	416.2	519.4	100.0	84.000	100
1950	422.4	360.7	101.9	127.1	108.5	402.5	511.0	98.4	83.601	85
1951	615.0	389.0	109.9	183.4	116.0	360.3	476.3	91.7	81.942	88
1952	694.8	435.6	123.1	198.2	124.3	454.5	578.8	111.4	96.735	94
1953	704.2	447.7	126.5	201.8	128.3	558.7	687.0	132.3	109.566	93
1954	797.3	500.8	141.5	219.8	138.1	689.6	827.7	159.4	129.063	91
1955	900.9	532.8	150.6	240.3	142.1	825.8	967.9	186.3	146.853	89
1956	1,027.6	546.6	154.5	256.1	136.2	808.4	944.6	181.9	145.398	86
1957	1,101.1	567.3	160.3	302.3	155.7	895.3	1,051.0	202.3	158.592	85
1958	1,231.4	595.7	168.4	360.0	174.2	901.9	1,076.1	207.2	163.572	92
1959	1,297.3	630.7	178.3	422.3	205.3	1,024.2	1,229.5	236.7	183.060	94

Sources: Column 1 — SVT, *Teollisuustilastoa.*
— *Tilastokatsauksia* No. 7/1960, p. 60.
2—3 — Column 1, above, and Table II.
4 — 1948—53: computed from *Teollisuustilastoa* on the assumption that the total wages of the newspapers' wage earners were 6.5 per cent of the total wages of wage earners in the printing industries. 1954—59: SVT, *Teollisuustilastoa; Tilastokatsauksia.*
5 — Column 4, above, and Table II.
6 — Columns 1 and 14 of Table II.
9 — Weights .27 and .57: *Lehtien postimaksukomitean mietintö* 1959, p. 55.
10 — Column 9 and Table II, column 1.

[1] Net national product.

Table IV. Observed values of variables[1] used in the study. 1949—1960. 1949 = 1.00 (Ref. pp. 27—30, above).

Calendar year	Q_t	A_{t-1}	C_t	Y_{t-1}	P_{t-1}	p_t	G_{t-1}	π_{t-1}	$R_{A_{t-1}}$	R_{C_t}	S_{t-1}	B_t
1949	1.00	1.07	1.00	.98	.78	1.00	.82	.99	1.06	1.00	.92	1.00
1950	1.17	1.00	1.07	1.00	1.00	.83	1.00	1.00	1.00	.89	1.00	1.00
1951	1.10	1.13	1.09	1.04	.83	.86	.85	.96	1.08	. 94	1.15	1.13
1952	1.21	1.18	1.14	1.14	.78	1.04	.88	.91	1.07	1.19	1.24	.93
1953	1.37	1.34	1.19	1.12	.89	1.08	.94	.94	1.26	1.29	1.35	.96
1954	1.62	1.40	1.22	1.13	.95	1.07	.93	.95	1.33	1.31	1.27	1.05
1955	1.86	1.67	1.24	1.22	.99	1.07	.91	. 94	1.57	1.33	1.41	1.07
1956	1.90	1.93	1.21	1.29	1.02	1.00	.89	.99	1.91	1.21	1.68	.99
	(2.05)[2]		(1.31)[2]							(1.31)[2]		
1957	2.06	1.85	1.28	1.29	.96	1.09	.86	1.04	1.92	1.40	1.71	1.03
				(1.31)[2]								
1958	1.98	1.85	1.25	1.28	.97	1.17	.85	1.08	2.00	1.46	1.58	.96
1959	2.13	1.90	1.21	1.28	1.02	1.27	.92	1.14	2.17	1.55	1.53	1.04
1960	2.33	2.03	1.26	1.35	1.06	1.21	.94	1.17	2.33	1.52	1.70	1.13

Sources: — Tables II and III.
The business cycle variable B_t has been computed by dividing the volume index of exports in 1949—59 by its trend values (see Table II, column 16). The equation is presented on page 30, above.

[1] Key to symbols:
Q_t = sales on the domestic market of Finnish-made newsprint, in kg. per inhabitant [3], during the year *t*
A_{t-1} = consumption of advertising space per inhabitant [3], during the year *t-1*
C_t = circulation; copies per inhabitant [3], in the year *t*
Y_{t-1} = real national product per inhabitant [3], in the year *t-1*
P_{t-1} = price of newsprint in Fmk./kg., deflated by the NNP price index, during the year *t-1*
p_t = mean annual subscription rate deflated by the NNP price index, in the year *t*
G_{t-1} = index of the cost level of the newspapers deflated by the NNP price index, in the year *t-1*
π_{t-1} = mean annual advertising space rate deflated by the NNP price index, in the year *t-1*
$R_{A_{t-1}}$ = newspapers' revenue from advertisements, deflated by the NNP price index, in the year *t-1*
R_{C_t} = revenue from sales of newspapers, deflated by the NNP price index, in the year *t*
S_{t-1} = volume of gross and retail sales in the year *t-1*
B_t = business cycle variable

[2] The values from which the influence of the General Strike (1956) has been eliminated, are shown in brackets.

[3] In this study, only inhabitants of 15 years of age or more are considered, see footnote on p. 37.

Table V. Basic values of variables given in Table VII. 1926—1938.

Calendar year	Sales of newsprint	Population of 15 years of age or over	Sales of newsprint per inhabitant [1]	Price of newsprint	Index of real national product	Real national product per inhabitant	Number of copies distributed by mail	Number of copies distributed by mail per inhabitant	Mean subscription rates for four Helsinki papers	Price index of net national product	Price of newsprint deflated by the NNP price index	Column 9 deflated by the NNP price index	Volume index of exports
	1	2	3	4	5	6	7	8	9	10	11	12	13
	1,000 kg.	1,000	kg.	Fmk./kg.	1938 = 100	1927 = 100	Millions	1927 = 100	Fmk.	1926 = 100	Fmk./kg.	Fmk.	1925 = 100
1926	9,837	2,309	4.26	3.09	64	93	168	92	172.4	100.00	3.09	172.4	103
1927	12,100	2,344	5.16	2.90	70	100	185	100	172.4	102.36	2.83	168.4	115
1928	12,714	2,377	5.35	2.75	72	101	201	107	173.3	105.19	2.61	164.7	114
1929	14,161	2,406	5.89	2.65	71	99	212	112	174.1	103.10	2.57	168.9	121
1930	13,299	2,432	5.47	2.66	70	96	217	113	174.1	96.26	2.76	180.9	109
1931	12,024	2,457	4.89	2.54	67	91	198	102	174.1	88.80	2.86	196.1	110
1932	10,546	2,483	4.25	2.54	68	92	182	93	174.1	87.47	2.90	199.0	116
1933	10,965	2,511	4.37	2.28	73	97	183	92	174.1	85.53	2.67	203.6	134
1934	12,285	2,533	4.85	2.05	82	108	191	96	174.1	86.08	2.38	202.3	145
1935	14,708	2,560	5.75	2.05	84	110	215	106	174.1	88.31	2.32	197.1	155
1936	16,299	2,593	6.29	2.05	90	116	237	116	174.1	91.78	2.23	189.7	173
1937	19,625	2,623	7.48	2.05	99	126	254	123	189.7	100.53	2.04	188.7	186
1938	21,781	2,654	8.21	2.04	100	127	287	137	199.1	103.77	1.97	191.9	160

Sources: Column 1—6 — See the corresponding sources for Table II.
7 — *Board of Posts and Telegraph*, and *Lehtien postimaksukomitean mietintö* 1959, p. 14.
9 — Table VI.
10 — *Bank of Finland Institute for Economic Research.*
13 — *Tilastokatsauksia.*

[1] See footnote, p. 37.

Table VI. Computation of the basic values of the variable p'_t (Table VII). Mean annual subscription rates for *Helsingin Sanomat, Hufvudstadsbladet, Suomen Sosialidemokraatti* and *Uusi Suomi*, 1926—1938.

Calendar year	Helsingin Sanomat	Hufvudstadsbladet	Suomen Sosialidemokraatti	Uusi Suomi	.52 × HS subscription rate	.14 × Hbl subscription rate	.08 × SSD subscription rate	.26 × US subscription rate	Weighted mean subscription rate	Column 9, deflated by the NNP price index
	Annual subscription rate, Fmk.									
	1	2	3	4	5	6	7	8	9	10
1926	165	230	144	165	85.8	32.2	11.5	42.9	172.4	172.4
1927	165	230	144	165	85.8	32.2	11.5	42.9	172.4	168.4
1928	165	230	155	165	85.8	32.2	12.4	42.9	173.3	164.7
1929	165	230	165	165	85.8	32.2	13.2	42.9	174.1	168.9
1930	165	230	165	165	85.8	32.2	13.2	42.9	174.1	180.9
1931	165	230	165	165	85.8	32.2	13.2	42.9	174.1	196.1
1932	165	230	165	165	85.8	32.2	13.2	42.9	174.1	199.0
1933	165	230	165	165	85.8	32.2	13.2	42.9	174.1	203.6
1934	165	230	165	165	85.8	32.2	13.2	42.9	174.1	202.3
1935	165	230	165	165	85.8	32.2	13.2	42.9	174.1	197.1
1936	165	230	165	165	85.8	32.2	13.2	42.9	174.1	189.7
1937	195	230	165	165	101.4	32.2	13.2	42.9	189.7	188.7
1938	195	230	185	195	101.4	32.2	14.8	50.7	199.1	191.9

Sources: Column 1—4 — The appropriate publishers.
5—8 — Weights obtained from a memorandum by KOIVULEHTO (1949, Appendices).
10 — NNP price index taken from Table V.

Table VII. Observed values of variables[1] used in the study. 1927—1938. 1927 = 1.00. (Ref. p. 54, above).

Calendar year	Q_t	Y_{t-1}	P_{t-1}	p'_t	R'_{C_t}	B_t
1927	1.00	.93	1.09	1.00	1.00	1.00
1928	1.04	1.00	1.00	.98	1.05	.93
1929	1.14	1.01	.92	1.00	1.12	.86
1930	1.06	.99	.91	1.07	1.21	.80
1931	.95	.96	.98	1.16	1.18	.77
1932	.82	.91	1.01	1.18	1.10	.77
1933	.85	.92	1.02	1.21	1.11	.85
1934	.94	.97	.94	1.20	1.15	.89
1935	1.11	1.08	.84	1.17	1.24	.91
1936	1.22	1.10	.82	1.13	1.31	.97
1937	1.45	1.16	.79	1.12	1.38	1.01
1938	1.59	1.26	.72	1.14	1.56	.84

Sources: — Tables V and VI.
— The variable B_t has been computed by dividing the volume index of exports in 1926—38 (see Table V, column 13) by its trend values. The equation is presented on page 55, above.
— The variable R'_{C_t} has been computed as the product of column 8 in Table V and p'_t.

[1] Key to symbols:
— Q_t, Y_{t-1}, P_{t-1}, and B_t: see footnote to Table IV.
— p'_t = mean annual subscription rate for *Helsingin Sanomat*, *Hufvudstadsbladet*, *Suomen Sosialidemokraatti* and *Uusi Suomi*. The weights used: 52, 14, 8 and 26, respectively (see p. 54, above).

Table VIII. Successive ratios of variables given in Table IV. ($\varDelta X = X_t/X_{t-1}$). (Ref. p. 42, above).

$\varDelta Q_t$	$\varDelta A_{t-1}$	$\varDelta C_t$	$\varDelta Y_{t-1}$	$\varDelta P_{t-1}$	$\varDelta p_t$	$\varDelta G_{t-1}$	$\varDelta \pi_{t-1}$	$\varDelta R_{A_{t-1}}$	$\varDelta R_{C_t}$	$\varDelta S_{t-1}$	$\varDelta B_t$
1.17	.93	1.07	1.02	1.28	.83	1.22	1.01	.94	.89	1.09	1.00
.94	1.13	1.02	1.04	.83	1.04	.85	.96	1.08	1.06	1.15	1.13
1.10	1.04	1.05	1.10	.94	1.21	1.04	.95	.99	1.27	1.08	.82
1.13	1.14	1.04	.98	1.14	1.04	1.07	1.03	1.18	1.08	1.09	1.03
1.18	1.04	1.03	1.01	1.07	.99	.99	1.01	1.06	1.02	.94	1.09
1.15	1.19	1.02	1.08	1.04	1.00	.98	.99	1.18	1.02	1.11	1.02
1.02	1.16	.98	1.06	1.03	.93	.98	1.05	1.22	.91	1.19	.93
1.08	.96	1.06	1.00	.94	1.09	.97	1.05	1.01	1.16	1.02	1.04
.96	1.00	.98	.99	1.01	1.07	.99	1.04	1.04	1.04	.92	.93
1.08	1.03	.97	1.00	1.05	1.09	1.08	1.06	1.08	1.06	.97	1.08

Source: — Table IV.

Table IX. Computation of mean subscription rate of newspapers (p_t) for 1960. (Ref. p. 58, above).

Newspaper	Amount of newsprint used		Annual subscription rate, Fmk.	
	1,000 kg.	Per cent	1957	1960
Helsingin Sanomat	6,048.0	33.7	3,100	4,000
Hufvudstadsbladet	1,761.4	9.8	3,100	4,000
Uusi Suomi	3,188.0	7.8	3,100	4,000
Suomen Sosialidemokraatti	842.2	4.7	2,700	3,400
Etelä-Suomen Sanomat	266.2	1.5	2,260	2,420
Savon Sanomat	654.7	3.6	2,300	2,400
Kaleva	319.4	1.8	2,210	2,200
Satakunnan Kansa	364.8	2.0	2,260	2,500
Aamulehti	1,300.0	7.2	2,360	2,650
Turun Sanomat	857.6	4.8	2,260	2,500
Uusi Aura	258.2	1.4	2,260	2,500
Ilkka	396.0	2.2	2,200	2,400
Vaasa	648.4	3.6	2,200	2,400
Palkkatyöläinen	497.4	2.8	500	500
Maakansa	202.2	1.1	2,360	2,400
Kansan Tahto	190.2	1.1	2,060	2,000
Kansan Lehti	164.4	.9	1,900	2,300
Total	17,959.1	100.0	.	.
Weighted mean	.	.	2,746.2	3,392.6

Sources: — *Posti- ja lennätinlaitos.*
— KOIVULEHTO (1949, Appendices).

Table X. Estimation statistics of models based on the economy of the newspapers. (Ref. pp. 37—41, above).

Number of model	16	17	18	19	20
To be explained	$\log Q_t$	$\log Q_t$	$\log Q_t$	$\log Q_t$	$\log Q_t$
Regression coefficient of variable	Estimates computed from original data				
1.000	.102 ± .201	.081 ± .022	.028 ± .018	.085 ± .017	.036 ± .022
$\log R_{A_{t-1}}$	.420 ± .132	.774 ± .129	.518 ± .070	.	.
$\log R_{C_t}$	.551 ± .174	.302 ± .208	.	.440 ± .143	.
$\log P_{t-1}$	.855 ± .235	.	.464 ± .157	.837 ± .174	.599 ± .189
$\log G_{t-1}$	.	.896 ± .373	.	.	.
$\log C_t$	.	.	1.345 ± .253	.	.916 ± .375
$\log A_{t-1}$	.	.	.	.574 ± .123	.670 ± .118
R	.982	.972	.991	.990	.987
s	.0228	.0286	.0158	.0175	.0198

Number of model	16a	18a	19a	20a
To be explained	$\log \Delta Q_t$	$\log \Delta Q_t$	$\log \Delta Q_t$	$\log \Delta Q_t$
Regression coefficient of variable	Estimates computed from first logarithmic differences			
1.000	.016 ± .145	.013 ± .015	.013 ± .013	.013 ± .013
$\log \Delta R_{A_{t-1}}$	.056 ± .253	.179 ± .265	.	.
$\log \Delta R_{C_t}$	.391 ± .252	.	.409 ± .242	.
$\log \Delta P_{t-1}$	.663 ± .218	.406 ± .166	.695 ± .222	.435 ± .169
$\log \Delta C_t$	.	1.199 ± .651	.	1.109 ± .589
$\log \Delta A_{t-1}$	.	.	.139 ± .254	.177 ± .249
R	.651	.695	.669	.698
s	.0262	.0248	.0257	.0247

The following tabulation gives the values of t (*cf.* p. 36, above) that are significant at a risk level of 10—.1 per cent with 6—7 degrees of freedom.

Level of risk / Degrees of freedom	10	5	1	.1
6	1.94	2.45	3.71	5.96
7	1.89	2.36	3.50	5.41

Table XI. Estimation statistics of models derived from the economy of the newspapers. (Ref. pp. 44—45, above).

Number of model	21	22	Number of model	21a	22a
To be explained	$\log Q_t$	$\log Q_t$	To be explained	$\log \Delta Q_t$	$\log \Delta Q_t$
Regression coefficient of variable	Estimates computed from original data		Regression coefficient of variable	Estimates computed from first logarithmic differences	
1.000	.070 ± .045	.109 ± .020	1.000	.021 ± .105	.014 ± .013
$\log S_{t-1}$	.648 ± .355	.487 ± .162	$\log \Delta S_{t-1}$	.100 ± .232	.129 ± .244
$\log C_t$	.972 ± .979	.	$\log \Delta C_t$	1.030 ± .586	.
$\log P_{t-1}$	.897 ± .352	1.054 ± .208	$\log \Delta P_{t-1}$	.397 ± .171	.418 ± .248
$\log R_{C_t}$	.	.621 ± .164	$\log \Delta R_{C_t}$	.	.648 ± .216
R	.948	.981	R	.679	.667
s	.0385	.0236	s	.0254	.0257

Table XII. Estimation statistics of models based on the economy of the newspapers' customers.[1] (Ref. pp. 46—48, 51—54, 58—59, above).

Number of model	23	26	23 with observations for 1956 adjusted	26 with observations for 1956 adjusted
To be explained	$\log Q_t$	$\log Q_t$	$\log Q_t$	$\log Q_t$
Regression coefficient of variable	Estimates from 1949—59 data			
1.000	.108 ± .016	.091 ± .013	.105 ± .014	.092 ± .012
$\log Y_{t-1}$	1.572 ± .198	1.322 ± .233	1.715 ± .171	1.439 ± .239
$\log P_{t-1}$	.948 ± .148	.900 ± .136	.977 ± .131	.927 ± .126
$\log p_t$	.510 ± .142	.	.374 ± .125	.
$\log B_t$	.500 ± .215	.449 ± .200	.378 ± .191	.396 ± .129
$\log R_{C_t}$	.	.470 ± .120	.	.380 ± .188
R	.992	.993	.994	.994
s	.0155	.0145	.0138	.0136
	Estimates from 1949—60 data			
1.000	..	..	.106 ± .016	.095 ± .013
$\log Y_{t-1}$	..	..	1.692 ± .203	1.395 ± .260
$\log P_{t-1}$	..	..	.962 ± .155	.924 ± .137
$\log p_t$	..	.	.322 ± .146	.
$\log B_t$	..	..	.188 ± .197	.240 ± .180
$\log R_{C_t}$	.	..	.	.388 ± .141
R	..	..	.992	.993
s	..	..	.0165	.0149

[1] Subscribers, advertisers and printing clients.

Table XIII. Estimation statistics of models based on the economy of the newspapers' customers.[1] Estimates obtained from first logarithmic differences. 1949—1959 data.
(Ref. pp. 50—54, above).

Number of model	23a	26a	23a with observations for 1956 adjusted	26a with observations for 1956 adjusted
To be explained	$\log \Delta Q_t$	$\log \Delta Q_t$	$\log \Delta Q_t$	$\log \Delta Q_t$
Regression coefficient of variable				
1.000	—.001 ± .013	—.005 ± .009	—.001 ± .010	—.003 ± .009
$\log \Delta Y_{t-1}$	1.392 ± .592	1.298 ± .389	1.560 ± .443	1.383 ± .361
$\log \Delta P_{t-1}$	.924 ± .234	.907 ± .142	.897 ± .181	.887 ± .143
$\log \Delta p_t$	.550 ± .258	·	.341 ± .198	·
$\log \Delta B_t$	.499 ± .247	.496 ± .162	.340 ± .184	.388 ± .165
$\log \Delta R_{c_t}$	·	.593 ± .154	·	.448 ± .194
R	.773	.898	.877	.907
s	.0219	.0152	.0170	.0149

[1] See footnote, p. 42.

Publications of the Society of Forestry in Finland:

ACTA FORESTALIA FENNICA. Contains scientific treatises dealing mainly with forestry in Finland and its foundations. The volumes, which appear at irregular intervals, generally contain several treatises.

SILVA FENNICA. Contains essays and short investigations mainly on forestry in Finland. Published at irregular intervals.

Die Veröffentlichungsreihen der Forstwissenschaftlichen Gesellschaft in Finnland:

ACTA FORESTALIA FENNICA. Enthalten wissenschaftliche Untersuchungen vorwiegend über die finnische Waldwirtschaft und ihre Grundlagen. Sie erscheinen in unregelmässigen Abständen in Bänden, von denen jeder im allgemeinen mehrere Untersuchungen enthält.

SILVA FENNICA. Diese Veröffentlichungsreihe enthält Aufsätze und kleinere Untersuchungen vorwiegend zur Waldwirtschaft Finnlands. Sie erscheint in zwangloser Folge.

Publications de la Société forestière de Finlande:

ACTA FORESTALIA FENNICA. Contient des études scientifiques principalement sur l'économie forestière en Finlande et sur ses bases. Paraît à intervalles irréguliers en volumes dont chacun contient en général plusieurs études.

SILVA FENNICA. Contient des articles et de petites études principalement sur l'économie forestière de Finlande. Paraît à intervalles irréguliers.

Helsinki, Unionink. 40. B.