ETLA
ELINKEINOELÄMÄN TUTKIMUSLAITOS
THE RESEARCH INSTITUTE OF THE FINNISH ECONOMY
Lönnrotinkatu 4 B, 00120 Helsinki 12, Finland, tel. 601322

Keskusteluaiheita
Discussion papers

Timo Teräsvirta

SHORT-TERM FORECASTING OF INDUSTRIAL PRODUCTION BY MEANS OF QUICK INDICATORS*

No. 109 23 June 1982

* During the course of the work I have benefitted from discussions with Åke Holmén, Jussi Karko and Pentti Vartia.

SHORT-TERM FORECASTING OF INDUSTRIAL PRODUCTION BY MEANS OF QUICK INDICATORS

by

Timo Teräsvirta

Abstract. This paper is an abridged version of a more detailed report available only in Finnish. It discusses building transfer function models with linear combinations of quick indicators as inputs for very short-term prediction of the monthly time series of the volume of industrial production in Finland. The number of input variables in the transfer function models is reduced by replacing the original indicators by their two first principal components. The prediction accuracy of the transfer function models is checked outside the sample and found superior to that of corresponding ARIMA models. It is also found that the prediction accuracy of the transfer function models compares rather favourably with the preliminary values of the volume of industrial production published by the Central Statistical Office. In 1979 and 1980 the final level of the volume is even approximated better by the models than by the series of preliminary values.

1. Introduction

This paper reports results on constructing a quantitative prediction model for quick forecasting of the present and the very near future of the monthly volume of industrial production in Finland using time series with short publication lags. The first preliminary value of the monthly volume of industrial production is available from the Central Statistical Office (C.S.O.) only after a lag of more than two months and, for a number of reasons, shortening this lag would be desirable.

This way of defining the task rules out structural models for the volume of industrial production, as at least some of the necessary time series are published no faster than the industrial production. The potentially useful time series naturally have to be related directly or indirectly to the industrial production, but they have to be published quickly. Models based on these variables which will be called indicators, will inavoidably be non-causal as the criteria for choosing the time series emphasize quick publication and bypass most economic theory.

The simplest models in this work are merely based on the past values of the volume of industrial production itself. They are used to check the prediction performance of more complicated models with genuine input variables. For the latter models to be useful they have to predict better than the univariate models, at least when the values of the input variables are known in advance.

The plan of the paper is as follows: In Section 2 the variables are briefly introduced. Section 3 contains some general remarks on transfer function models in view of the present problem while the estimated models are presented in Section 4. The prediction performance of the models is discussed in Section 5, and the last section contains a brief account of how to use the models in practice.

2. Indicator variables

As mentioned previously, the monthly indicator variables have been chosen with two things in mind. They have to be related to the volume of industrial production and published quickly. The following six monthly time series will be considered:

- the electricity consumption (X_{1t})[1)]
- index of advertisement space in newspapers (X_{2t})
- number of State Railways freight cars loaded (X_{3t})
- number of vacancies (X_{4t})
- number of unemployed (X_{5t})
- volume of exports (X_{6t})

The first, third and last series are directly connected with the industrial output while the remaining ones are more loosely related to it. The logarithmed and differenced series together with the logarithmic volume of industrial production (y_t) and its twelve month difference are in Figures 1 - 8. A more thorough discussion of these series and where they are published is in Teräsvirta (1982). All of them are available at a few weeks' notice and their first published values are final. Some at least potentially useful series cannot be considered because they are only published quarterly. The list is quite different from the set of variables in Neftci (1979) who discusses the prediction of U.S. industrial production by means of leading and coincident indicators. It should be mentioned that the first values of the monthly volume of Finnish industrial production are preliminary and subject to change, and the final values will eventually appear more than a year later. When the base year of the index is changed, which happens at five years' intervals, even some of these final monthly values may change.

3. Transfer function models

The models in this paper are linear and based on the logarithms of the original series. The family of models to be applied can be written as

$$D(z)y_t = \mu + \sum_{i=1}^{k} \delta_i(z)^{-1}\omega_i(z)D(z)x_{it} + \phi(z)^{-1}\Phi(z^s)^{-1}\theta(z)\Theta(z^s)a_t \tag{3.1}$$

where $a_t \sim nid(0,\sigma^2)$, z is the lag operator, $zx_t = x_{t-1}$, and D(z) is an operator whose roots are on the unit circle. Furthermore,

$$\omega_i(z) = \sum_{j=0}^{u_i} \omega_{ij}z^j , \quad \delta_i(z) = 1 - \sum_{j=1}^{r_i} \delta_{ij}z^j , \quad i = 1,\dots,k$$

$$\phi(z) = 1 - \sum_{j=1}^{p} \phi_j z^j , \quad \theta(z) = 1 - \sum_{j=1}^{q} \theta_j z^j$$

$$\Phi(z^s) = 1 - \sum_{j=1}^{P} \Phi_j z^{js} , \quad \Theta(z^s) = 1 - \sum_{j=1}^{Q} \Theta_j z^{js}$$

and all the roots of the above polynomials are outside the unit circle. Using the terminology of Box and Jenkins (1970), (3.1) is a single output, multiple input transfer function model with a multiplicative stationary ARMA(p,q) x $(P,Q)_s$ error process. The orders of the polynomials are not assumed known in advance, neither is s fixed to equal 12.

In view of the present application, (3.1) will potentially contain a relatively large number of parameters. The number of indicators is six and the appearance of lags is not excluded. As the model is non-causal, it is hardly realistic to expect the errors to be white noise. This means that an adequate description of error dynamics also requires parameters. On the other hand, some indicators are only available from the year 1969 onwards, so that the number of observations would not be particularly high with respect to the number of parameters and expected multicollinearity in the data. Model (3.1) is also somewhat impractical when forecasting is concerned. As long as predicting the present is the issue, there are no problems. However, if the model is used for genuine short-term forecasting, then the relatively large number of indicators is a drawback, because each indicator has to be forecast in order to obtain a prediction of the output variable. For all these reasons, reducing the dimension of the model might be considered desirable, and this can be achieved through a proper transformation of variables.

4. ARIMA models and principal components

Before turning to the reduction of the model we report results for specifying and estimating ARIMA models for the volume of industrial production. All the time series in this study are seasonally unadjusted. The monthly observations in the sample consisted of the years 1969-1979 and the year 1980, the last one for which almost final values of industrial production of the volume existed at the moment, was used for prediction. Details of the specification are not discussed here, but the maximum likelihood estimation[2)] of the specified model yields

$$\nabla\nabla_{12}y_t = (1 + \underset{(0.10)}{0.11}z^2 + \underset{(0.10)}{0.06}z^4)^{-1}(1 - \underset{(0.08)}{0.66}z)\hat{a}_t \tag{4.1}$$

$s = 0.0481$, $Z_1(9) = 5.2(0.186)$,

$\text{med}(e^*) = 0.0276$, $\text{med}|e^*| = 0.0456$, $\text{rmse}(e^*) = 0.0641$

where $\nabla_h = 1 - z^h$, s is the standard deviation of the residuals, Z_1 is the Box-Pierce test statistic and the figure in parentheses is the corresponding c.d.f. value, med(e*) is the median of the prediction errors of the monthly values of 1980, med|e*| is the median of the corresponding absolute prediction errors (MAPE), and rmse(e*) is the root mean square error (RMSE) for the same period. The size of the RMSE is largely due to the big prediction error in July 1980. The year 1980 was a boom year, and therefore the paper and pulp industry were working on full capacity even in July which is a traditional holiday month. The volume of industrial production is then at its lowest so that a large relative prediction error in July is not crucial. Judging from the median, model (4.1) somewhat underestimated the trend of the volume of industrial production that year.

Proceeding now to transfer function models, a reduction in the dimension of the model was achieved by a linear transformation of variables. The two principal components containing most of the standardised total variation of the differenced original indicators were chosen to be the new input variables. For a discussion of principal components, see e.g. Anderson (1958). Omitting the remaining principal components from the model is equivalent to setting linear restrictions between the parameters in (3.1). There is no theory saying in which variable space the transformation should be performed. Both twelve month differences (∇_{12}) and differenced twelve month differences ($\nabla\nabla_{12}$) were experimented with, but the results differed relatively little, cf. Teräsvirta (1982). The variable space did not contain lags of the indicators, but then lags of the principal components were allowed for in transfer function models.

Table 1 displays the eigenvectors based on twelve month differences. The two first principal components p_{1t} and p_{2t} contain almost 78 p.c. of the standardised total variation of the data. The first principal component has a very clear-cut interpretation: the signs of the coefficients of the indicators agree with the signs of their correlations with the industrial production. The second eigenvector divides the indicators into two sets. The first set consists of the three direct indicators, cf. Section 2, with the same sign as in the first eigenvector while the remaining indirect indicators form a second set. The details in the specification of the lag structure of this and other transfer function models are discussed in Teräsvirta (1982).

The specification and subsequent estimation gave the following model

$$\nabla\nabla_{12}y_t = \underset{(0.004)}{(0.023} + \underset{(0.004)}{0.006z^3)}\nabla p_{1t} + \underset{(0.004)}{(0.012} - \underset{(0.004)}{0.007z^3} + \underset{(0.003)}{0.006z^6)}\nabla p_{2t}$$

$$+ (1 + \underset{(0.10)}{0.13}z^4)^{-1}(1 - \underset{(0.04)}{0.93}z)\hat{a}_t \qquad (4.2)$$

$s = 0.0362, \ Z_1(10) = 4.0(0.05)$

$\text{med}(e^*) = -0.0029, \ \text{med}|e^*| = 0.0335, \ \text{rmse}(e^*) = 0.0465.$

Since the model (4.2) is not causal there is no convincing interpretation for the estimated lag structure. It could be mentioned, however, that the autocorrelation function of $\nabla_{12}y_t$ already shows signs of periodic variation of three months in the series, so that the corresponding lags in the principal components obviously reflect this situation. The residual standard deviation of (4.2) is definitely lower than that of (4.1), indicating that the transfer function has explanatory power. Nevertheless, only forecasting is a true test for (4.2), and that will be the topic of the next section.

5. Forecasts

Table 2 contains some prediction statistics for ARIMA and transfer function models (4.1) and (4.2). It is immediately clear from the table that the indicators are useful in forecasting the volume of industrial production. The forecasts are not one-step-ahead predictions, but the whole year has been predicted at one time. Perhaps the most conspicuous feature when the transfer function model is concerned is the improvement in forecasting the level of the volume. The ARIMA models underestimate the development in 1979 and 1980 which have been years of rapid growth while the transfer function models predict the trend very well. The most critical month for the models is July for the reason discussed above, but even there the transfer function models are superior to the ARIMA models. It can be mentioned that any of the six single indicators is already an improvement as compared to (4.1). Teräsvirta (1982) reports some results on single indicator transfer function models showing this, but the models with several indicators like (4.2) perform even better.

In order to obtain an idea of the relative size of the statistics in Table 2 we have computed their values also for the preliminary values of the volume of industrial production and included them in the same table. The preliminary values in question are the first values published by the C.S.O. and are subject to later revisions. It can be seen by comparing the two medians that the preliminary data have contained systematic downward bias both in 1979 and 1980. For 1979 the bias has been so large that the transfer function model in fact outperforms the preliminary values of the C.S.O. in terms of MAPE. In theory, this should not be so, as the preliminary values are based on a larger set of information than the indicator forecasts. Indeed, for 1980 the preliminary values are more accurate. Since they already contain direct information about the production, the final July values in particular are a lot better approximated by the preliminary values than by the model (4.2) or its equivalent for 1979. As a whole, however, the performance of the transfer function model may be deemed quite satisfactory.

An attractive observed feature of the transfer function models is the absence of systematic bias in prediction, see also Teräsvirta (1982). Since the predictions by the model play a rôle in forecasting central macroeconomic variables at the Research Institute of the Finnish Economy, this is important.

6. Forecasting in practice

In order to forecast with a model of type (4.2), the following steps need to be taken:

(i) The indicators and, subsequently, the two principal components are updated.

(ii) If forecasting is extended from the present into the future, the necessary values of principal components are predicted using ARIMA models. The principal components are uncorrelated only at lag zero, but the estimated cross correlations at other lags suggest that the loss of efficiency for not using a vector ARIMA model but rather two separate ARIMA models is not substantial.

(iii) The volume of industrial production is predicted using the transfer function model with the two principal components as inputs.

This simple routine requires neither a respecification of the model nor a re-estimation of its parameters each time new observations become available. It is easy to repeat whenever desired and, which is important, at a low cost. The transfer function model has so far proved accurate enough to be useful in practical work.

References

Anderson, T.W. (1958). *An introduction to multivariate statistical analysis.* New York: John Wiley.

Box, G.E.P. ja G.M. Jenkins (1970). *Time series analysis, forecasting and control.* San Francisco: Holden-Day.

Mellin, I., ed. (1980). SURVO 76, time series programs. University of Helsinki, Department of Statistics, mimeo.

Neftci, S.N. (1979). Lead-lag relations, exogeneity and prediction of economic time series. *Econometrica 47*, 101-113.

Teräsvirta, T. (1982). *Teollisuustuotannon volyymin lyhyen ajan ennustaminen osoitinmuuttujien avulla.* Helsinki: Research Institute of the Finnish Economy.

Footnotes

1) The capital letters refer to original variables while the lowercase letters represent logarithms of the original variables.

2) We have used the maximum likelihood estimation routine in the SURVO 76 statistical system, see Mellin (1980).

Table 1. The eigenvectors of the logarithmic twelve month differences of the indicators (from the correlation matrix) and the relative shares of the principal components of standardised total variation

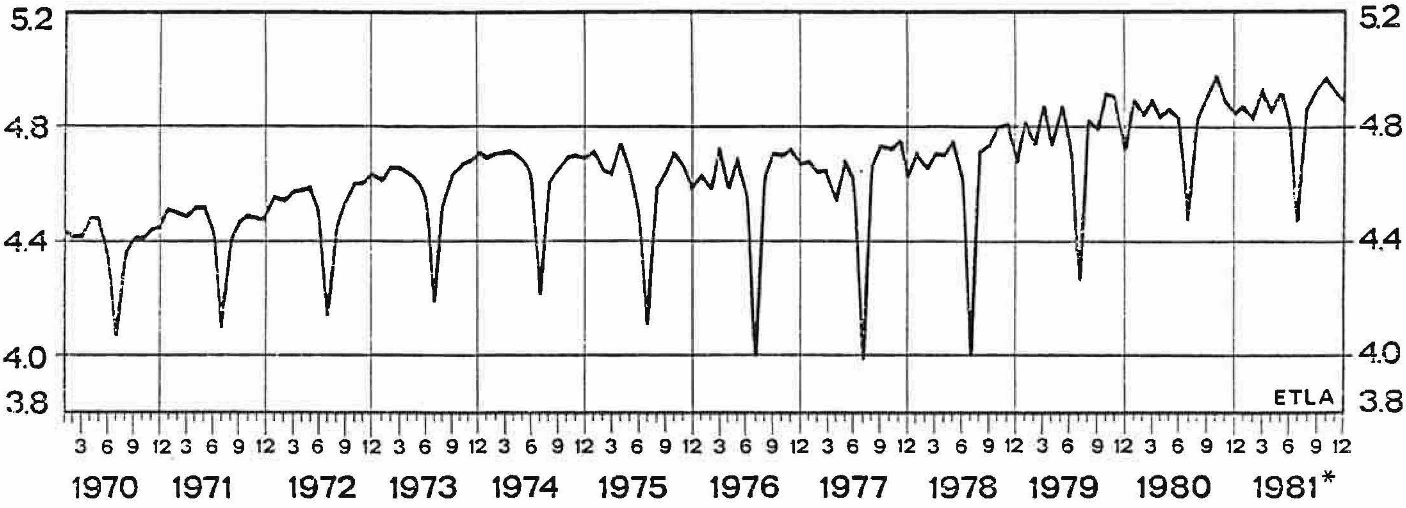
Variable	Eigenvector					
	1	2	3	4	5	6
$\nabla_{12}x_{1t}$	0.586	0.568	0.568	-0.050	-0.070	-0.040
$\nabla_{12}x_{2t}$	0.747	-0.421	0.102	0.407	0.295	0.011
$\nabla_{12}x_{3t}$	0.734	0.437	-0.238	-0.292	0.352	-0.032
$\nabla_{12}x_{4t}$	0.903	-0.260	-0.047	-0.129	-0.199	0.239
$\nabla_{12}x_{5t}$	-0.785	0.514	0.098	0.082	0.230	0.220
$\nabla_{12}x_{6t}$	0.379	0.787	-0.302	0.320	-0.204	-0.011
Relative share of total variation	0.503	0.274	0.082	0.063	0.059	0.018

Table 2. Statistics for prediction accuracy of models (4.1) and (4.2) and preliminary data: (a) RMSE, (b) MAPE, (c) median of prediction errors, (d) prediction error of July

Months forecast	statistic	Model*		Preliminary data
		(4.1)	(4.2)	
1980I-XII	(a)	0.0641	0.0465	0.0329
	(b)	0.0456	0.0335	0.0242
	(c)	0.0276	-0.0029	0.0242
	(d)	0.155	0.107	0.0135
1979I-XII	(a)	0.0833	0.0479	0.0408
	(b)	0.0475	0.0234	0.0364
	(c)	0.0475	-0.0082	0.0364
	(d)	0.223	0.136	0.0336

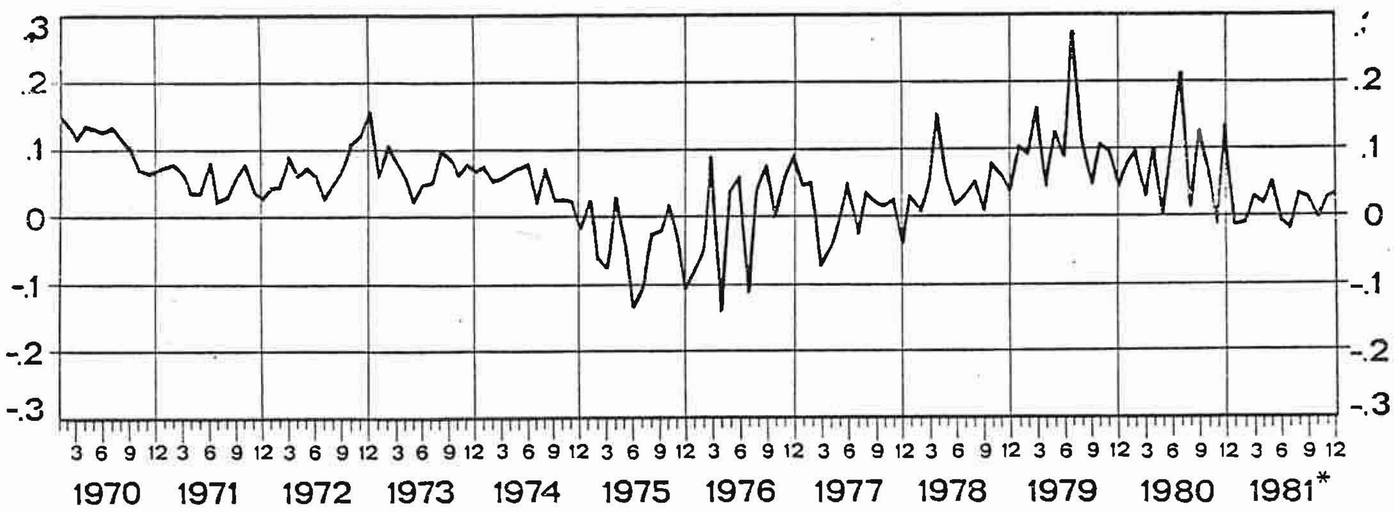
* Months 1979I-XII have been forecast by models corresponding to (4.1) and (4.2) whose estimation period has extended till December 1978 only. The estimated models are reported in Teräsvirta (1982).

FIGURE 1. THE LOGARITHMIC VOLUME OF INDUSTRIAL PRODUCTION, 1970 I - 1981 XII



* THE VALUES FOR 1981 ARE PRELIMINARY.

FIGURE 2. TWELVE MONTH DIFFERENCES OF THE LOGARITHMIC VOLUME OF INDUSTRIAL PRODUCTION, 1970 I - 1981 XII



* THE VALUES FOR 1981 ARE BASED ON PRELIMINARY VALUES.

FIGURE 3. LOGARITHMIC TWELVE MONTH DIFFERENCES OF ELECTRICITY CONSUMPTION (CONTINUOUS LINE) AND INDUSTRIAL PRODUCTION (DOTTED LINE), 1970 I - 1981 XII

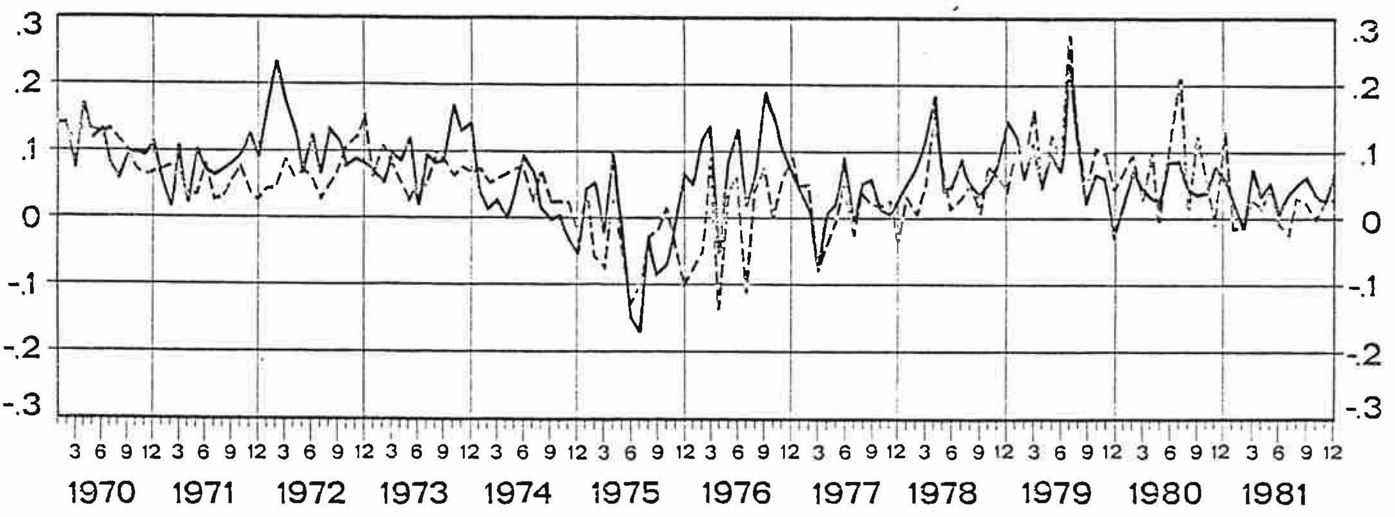


FIGURE 4. LOGARITHMIC TWELVE MONTH DIFFERENCES OF THE STATE RAILWAYS FREIGHT CARS LOADED (CONTINUOUS LINE) AND INDUSTRIAL PRODUCTION (DOTTED LINE), 1970 I - 1981 XII

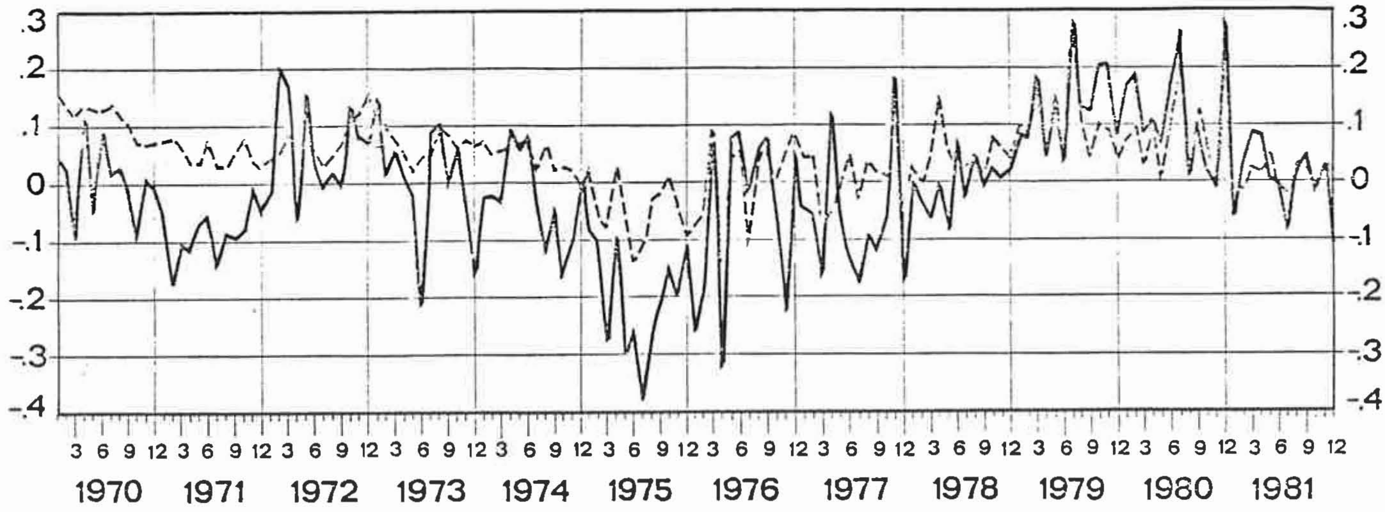


FIGURE 5. LOGARITHMIC TWELVE MONTH DIFFERENCES OF THE INDEX OF ADVERTISEMENT SPACE IN NEWSPAPERS (CONTINUOUS LINE) AND INDUSTRIAL PRODUCTION (DOTTED LINE), 1970 I - 1981 XII

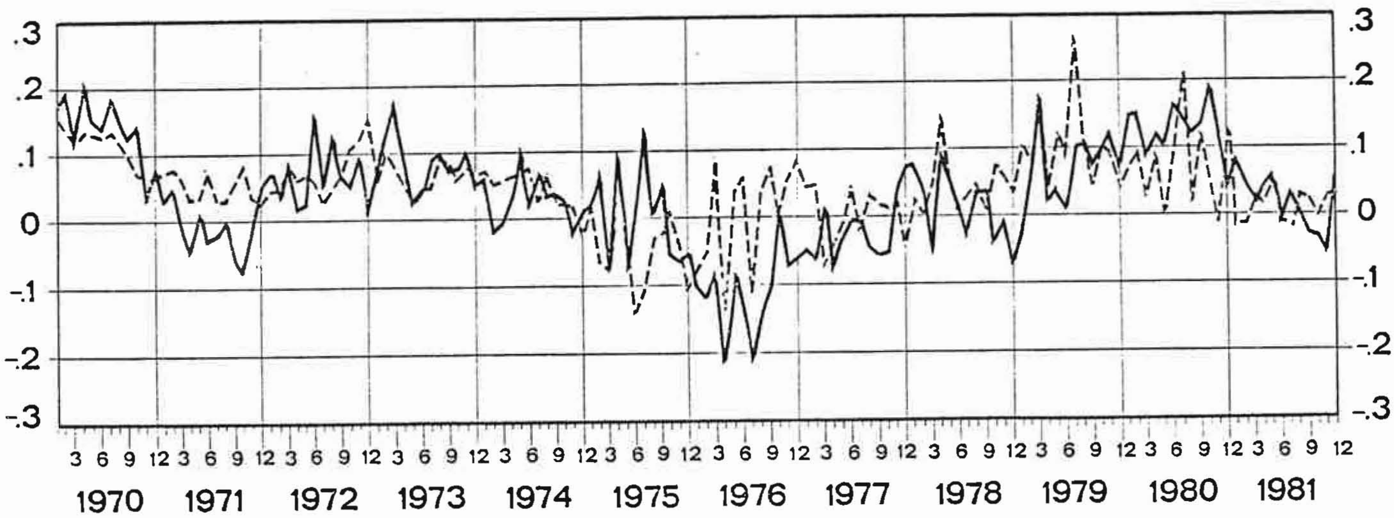
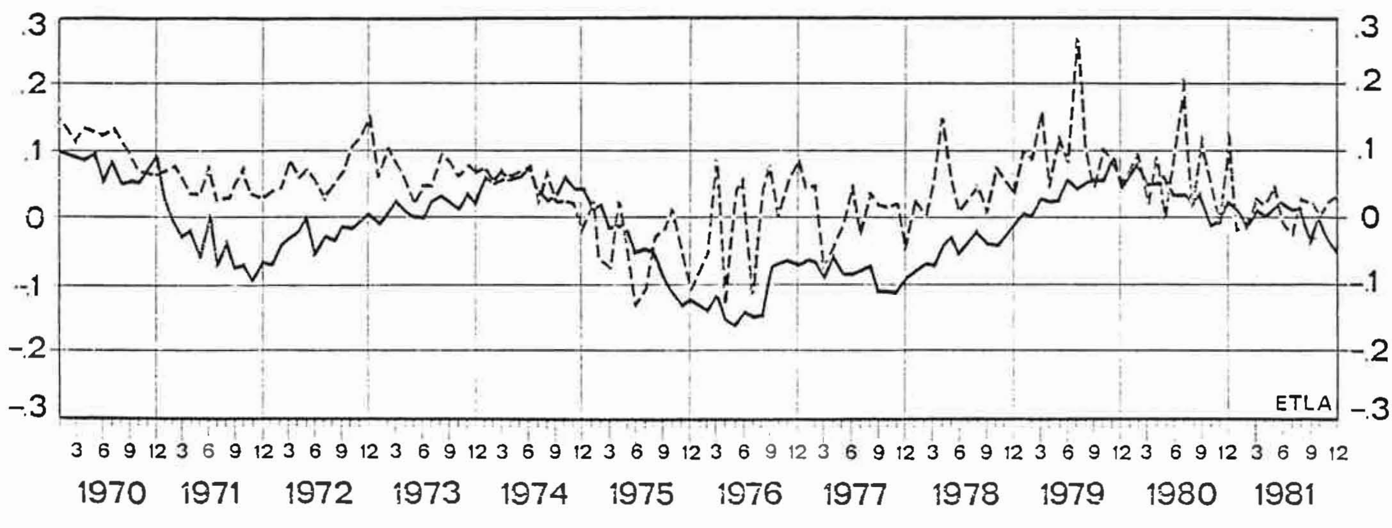
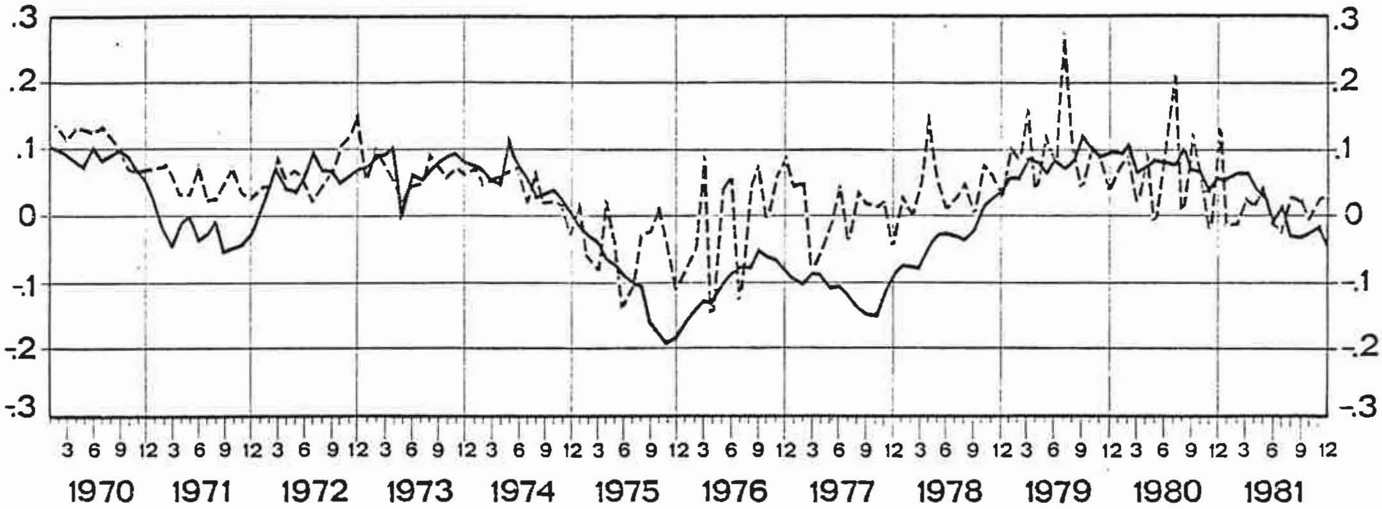


FIGURE 6. LOGARITHMIC TWELVE MONTH DIFFERENCES OF THE NUMBER OF UNEMPLOYED (CONTINUOUS LINE)* AND INDUSTRIAL PRODUCTION (DOTTED LINE), 1970 I - 1981 XII



* THE UNEMPLOYMENT SERIES HAS BEEN RESCALED BY CHANGING THE SIGN AND DIVIDING ITS VALUES BY FIVE.

FIGURE 7. LOGARITHMIC TWELVE MONTH DIFFERENCES OF THE NUMBER OF VACANCIES (CONTINUOUS LINE)* AND INDUSTRIAL PRODUCTION (DOTTED LINE), 1970 I - 1981 XII



* THE VACANCIES SERIES HAS BEEN RESCALED BY DIVIDING ITS VALUES BY FIVE.

FIGURE 8. LOGARITHMIC TWELVE MONTH DIFFERENCES OF THE VOLUME OF EXPORTS (CONTINUOUS LINE) AND INDUSTRIAL PRODUCTION (DOTTED LINE), 1970 I - 1981 XII