EGON SMERAL

Forecasting International Tourism Using ARIMA Models

61

EGON SMERAL

Forecasting International Tourism Using ARIMA Models

April 1993

WIFO Working Papers, 1993, (61)

Dr. Egon Smeral
Austrian Institute of
Economic Research
P.O. Box 91
A-1103 Vienna
Austria
Phone: 0043 222 78 26 01 219
Fax: 0043 222 78 93 86
E-Mail: SMERAL@WIFOSV.WSR.AC.AT

April 1993

FORECASTING INTERNATIONAL TOURISM USING ARIMA MODELS

Paper presented at the thirteenth Annual International Symposium on Forecasting, June 9-12, 1993, Pittsburgh, Pennsylvania, USA.

1. Introduction

Forecasting international tourism in a causal framework through the use of a structural model can be difficult or sometimes even impossible. This might be the case when, for example, data are not available for the exogeneous variables of an econometric model or for a set of regression equations which are believed to affect the endogeneous variables. Should the data be available, the estimation of regression equations for the set of endogeneous variables could result in standard errors that are so large as to make most of the estimated coefficients insignificant and the standard error of forecast unacceptably large. Thus even when one is able to estimate statistically significant regression equations for a set of dependent variables, the results may not be useful for forecasting purposes. To obtain a forecast for a dependent variable from a regression equation, explanatory variables that are not lagged must be themselves forecasted; this may be more difficult than forecasting the dependent variable itself. The standard error of forecast for the dependent variable with future values of the known explanatory variables may be small. However, when the future values of the explanatory variables are unknown, their forecast errors may be so large as to make the total forecast error for the dependent variable in question too large to be acceptable.

Contrary to a causal oriented forecasting system, a time series model accounts for patterns in the past movements of a variable, using that information in turn to predict its future movements. In this paper, time series models known as ARIMA models are used for forecasting international outbound tourism. Although the ARIMA model is frequently used in the wide field of practical economic forecasting, its application to tourism forecasting is almost non-existent.

The abbreviation ARIMA stands for "autoregressive integrated moving average". Integrated (I) refers to the "differencing" of the data series. The autoregression (AR), differencing (I) and moving average (MA) portion constitute the three numbers following ARIMA. The term integrated is used when the differencing is performed to achieve stationarity: the stationary series must be integrated (undifferenced) to recover the original data. For an ARIMA (p,d,q) model, the order is given by the three letters p, d and q. The order of the autoregressive component is p, the order of differencing needed to achieve stationarity is d, and the order of the moving average element is q.

ARIMA models have been studied extensively by George Box and Gwilym Jenkins. Their names have frequently been used synonymously with general ARIMA processes applied to time series analyses, forecasting and control.

2. Some properties of ARIMA models

ARIMA time series models are designed for stationary time series. A stationary time series is one whose basic statistical properties, such as the mean or variance, remain constant over time. Periodic variations and systematic changes in the mean and variance must be first identified and then removed in order to build ARIMA models. A time series is nonstationary when a pattern is present. In a nonstationary series the autocorrelation coefficents are typically statistically different from zero for the first several time lags, and only gradually drop to zero or show a spurious pattern as the number of time periods increases.

For the purpose of forecasting, the analyst must first identify the appropriate ARIMA model and transform and/or difference the data to produce stationarity (most time series encountered in some forecasting applications are not stationary series).

Achieving stationarity is reduced to the rather mechanical task of transforming time series until the autocorrelations drop to zero with one, two, or three time lags. Often (after transforming and/or differencing) the resulting series may require only an autoregressive component p or a moving average component q.

Autoregressive (AR) models express the time series variable Y_t as a linear function of some number of actual past values of Y_t. In general, in the autoregressive process of order p the current observation Y_t is generated by a weighted average of past observations going back p periods, together with a random disturbance et in the current period t.

$$Yt = c + a_1Y_{t-1} + a_2Y_{t-2} + ... + a_pY_{t-p} + e_t$$

$a_1 ..... a_{1-p}$ are model parameters; c is a constant term which relates to the mean u of the stochastic process.

If the autoregressive process is stationary, than its mean u must be invariant with respect to time, $E(Y_t) = E(Y_{t-1}) = E(Y_{t-2}) = ... = u$. The mean u is thus given by

$u = a_1u + a_2u + ...+ a_pu + c$ or

$u = c/(1 - a_1 - a_2 - ...-a_p)$

If the process is stationary, the mean u must be finite. If this were not the case, the process would drift farther and farther away from any fixed reference point and could not be stationary. If u is finite, it is necessary that

$a_1 + a_2 +......+a_p < 1$

These conditions are not sufficient to ensure stationarity, since there are other conditions necessary if the AR(p) process is to be stationary. I will discuss these additional conditions for special application cases in more detail below.

If AR(p) models cannot isolate certain data patterns when p is small, an alternative model - the (MA) moving average model - may be used to isolate the pattern when AR(p) models fail. In the moving average process of order q, each observation Yt is generated by a weighted average of random disturbances going back q periods.

$Y_t = u - b_1e_{t-1} - b_2e_{t-2} -....- b_qe_{t-q} + e_t$

The parameters $b_1.....b_q$ of the MA(q) model may be positive or negative.

Many stationary random processes cannot be modeled as either purely moving average or as purely autoregressive, as they have qualities of both types of processes. The ARMA (p,q) process I have denoted as

$$Y_t = c + a_1Y_{t-1} + a_2Y_{t-2} + \ldots + a_pY_{t-p} + e_t - \\ - b_1e_{t-1} - b_2e_{t-2} - \ldots - b_qe_{t-q}$$

For forecasting international tourism I only consider five special AR(I)MA models:

(1) AR(I)MA (1,0,0): $Y_t = c + a_1Y_{t-1} + e_t$

For stationarity to exist, the ARIMA (1,0,0) model requires that $-1<a_1<1$.

(2) AR(I)MA (2,0,0): $Y_t = c + a_1Y_{t-1} + a_2Y_{t-2} + e_t$

For a second-order autoregressive model, stationarity is assumed if

$$a_1 + a_2 < 1$$
$$a_2 - a_1 < 1$$
$$-1 < a_2 < 1$$

(3) AR(I)MA (0,0,1): $Y_t = u + e_t - b_1e_{t-1}$

With the models (1) and (2) I discussed the property of stationarity in a time series model. In modeling MA processes another property called invertibility is required in describing and forecasting a time series. It can be shown that, for any value of the parameter b_1, the ARIMA

(0,0,1) model describes the behavior of a stationary time series. Thus there are no conditions that must be imposed on b1 to make the model described stationary. However, certain conditions are imposed on the parameters of the MA models to insure invertibility. Invertibility refers to the possibility of inverting an MA model and expressing it as an AR model of infinite order; an AR model of infinite order has an infinite number of autoregressive coefficients. For an ARIMA (0,0,1) model, the invertibility condition is

$-1<b_1<1.$

(4) AR(I)MA (0,0,2): $Y_t = u + e_t - b_1e_{t-1} - b_2e_{t-2}$

In this type of model, invertibility is given if

$b_1 + b_2 < 1$
$b_2 - b_1 < 1$
$-1 < b_2 <1$

(5) AR(I)MA (1,0,1): $Y_t = c + a_1Y_{t-1} + e_t - b_1e_{t-1}$

It can be shown that the autoregressive moving-average model is stationary if

$-1 < a_1 < 1$

and is invertible if

$-1 < b_1 < 1$

3. Estimation of selected ARIMA models for international tourism

The models (1)-(5) were applied to yearly tourism import data of selected countries for the period 1975 - 1992. AR(I)MA models with lags not larger than 2 are very common for estimation procedures based on yearly data sets. The use of quarterly data was not possible as the calender effects (i.e., changes in vacation patterns, different easter or christmas holidays) would result in a strong bias in the estimation results; the lack of data availability is another reason why the estimation based on a yearly data set with only 18 observations for each variable is a feasilble - though not theoretically perfect- solution: it is a contribution to the discussion of forecasting methods in tourism.

The data source is the balance of payments statistics of the IMF. The import values are calculated for constant prices and exchange rates. In order to achieve stationarity of the time series, only the absolute deviations ($=TD_t$) from a quadratic trend ($Y_t = k + z_1 t + z_2 t_2$) are considered. An iterative nonlinear least squares procedure is applied to the parameter estimates of the models (1)-(5).

The identification of appropriate time series models for the sophisticated Box-Jenkins estimation was not possible, as there were by far too few observations be able to apply the Box-Jenkins approach (as a rule, at least 50 observations are required for Box-Jenkins estimation; even more observations are recommended for a seasonal model). The basis of their sophisticated approach consists of three phases: identification of the stationary time series model, estimation and testing, and application.

The estimation results for the parameters of the chosen models for outbound traveling from selected countries as well as Europe in total are summarized below.

Western Germany:

ARIMA (2,0,0); SEE = 0.405256

$$TD_t = -0.03600 + 1.02334*TD_{t-1} - 0.59219*TD_{t-2}$$

ARIMA (1,0,1); SEE = 0.430332

$$TD_t = -0.09883 + 0.44794*TD_{t-1} + e_t + 0.65327*e_{t-1}$$

United Kingdom:

ARIMA (2,0,0); SEE = 0.363265

$$TD_t = 0.05668 + 0.93640*TD_{t-1} - 0.73248*TD_{t-2}$$

Austria:

ARIMA (0,0,1); SEE = 0.111336

$TD_t = -0.00667 + e_t + 0.58830*e_{t-1}$

(for export models see also the annex)

France:

ARIMA (0,0,1); SEE = 0.286719

$TD_t = -0.07524 + e_t + 0.78680*e_{t-1}$

Netherlands:

ARIMA (0,0,2); SEE = 0.228025

$TD_t = -0.05080 + e_t + 0.62433*e_{t-1} + 0.46592*e_{t-2}$

Western Europe:

ARIMA (2,0,0); SEE = 1.03846

$TD_t = -0.25065 + 1.41088TD_{t-1} - 0.83018*TD_{t-2}$

ARIMA (1,0,1); SEE = 1.2027

$TD_t = -0.36306 + 0.68521*TD_{t-1} + e_t + 0.70193*e_{t-1}$

Canada:

ARIMA (0,0,2); SEE = 0.492145

$TD_t = -0.30969 + e_t + 1.01179*e_{t-1} + 0.72548*e_{t-2}$

Australia:

ARIMA (0,0,1); SEE = 0.195646

$TD_t = -0.00765 + e_t + 0.48350*e_{t-1}$

Japan:

ARIMA (2,0,0); SEE = 0.815545

$TD_t = 0.00519 + 0.88210*Y_{t-1} - 0.70523*Y_{t-2}$

The estimated model parameters fulfill the necessary stationarity and/or invertibility conditions; based on t-values, all estimated parameters are significant.

4. Forecast results

After parameter estimations of the models (1)-(5) forecasts until 1995 were completed.

The computation of the forecast can be done recursively using the estimated models. This involves first computing a forecast for one period ahead, then using this forecast to compute a forecast two periods ahead, and continuing in this manner until the chosen forecast time horizon has been reached.

To compute the forecast $Y_T(1)$, the start is set by computing the one-period forecast of TD_t, $TD_T(1)$. For that reason, we start with an ARIMA (p,d,q) equation that is one time period modified:

$$TD_{T+1} = a_1 TD_T + \ldots + a_p TD_{T-p+1} + \\ + e_{T+1} - b_1 e_T - \ldots - b_q e_{T-q+1} + c$$

The calculation of the forecast $TD_T(1)$ is done by taking the conditional expected value of TD_{T+1} in the last equation:

$$TD_T(1) = E(TD_{T+1}|TD_T,\ldots) = a_1TD_T +\ldots+ a_pTD_{T-p+1} - b_1e_T - \ldots - b_qe_{T-q+1} + c$$

where e_T, e_{T-1}, etc. are here observed residuals; the expected value of e_{T+1} is 0. We can obtain the two period forecast $TD_T(2)$:

$$TD_T(2) = E(TD_{T+2}|TD_T,\ldots) = a_1TD_T(1) + a_2TD_T +\ldots + a_pTD_{T-p+2} - b_2e_T - \ldots - b_qe_{T-q+2} + c$$

The two period forecast is then used to calculate the three period forecast, and so on, until the l-period forecast $TD_T(l)$ is reached:

$$TD_T(l) = a_1TD_{T(l-1)} + \ldots + a_1TD_T + \ldots a_pTD_{T-p+l} - b_le_T - \ldots - b_qe_{T-q+l} + c$$

Once the transformed series TD_t has been forecasted, a forecast can be obtained for the original series Y_t simply by adding the forecasted trend deviation to the values of the trend forecasts.

The forecast results are summarized in graph 1-9 and in table 1.

Based on % changes in the original import data, an analysis of the results shows us a recovery for the outbound traveling of western Europe in the year 1994 and 1995. However, the forecast results for the year 1993 lead one to expect a stagnation according to the ARIMA (2,0,0) model and a modest growth according to the ARIMA (1,0,1) model. Of the European countries selected, the growth expectations for Western Germany and the Netherlands are very disappointing: Western Germany could expect a modest growth rate in its tourism imports, the expenditures of the Netherlands will decrease in absolute terms. In all the other European countries, such as the UK, France and Austria, the recovery will be to a greater or lesser degree strong.

From the oversea countries, the tourism imports of Japan will expand very strongly. The growth rates for the tourism imports of Australia will be by far more moderate. Canadian expenditures abroad seem to expand more weakly after 1993 and will fall below the quadratic trend in the years 1994 and 1995.

In a broad sense, the forecasting results of the time series model fit partly to the possible results of a causal oriented forecasting system, as in many cases the present and expected patterns of the business cycles are reflected in the forecast results. But it is questionable if it would be possible to construct an econometric model or to estimate a multiple regression equation which would have forcasted the turning points of the outbound travel as well, given the forecast information available for the exogeneous variables.

Summary

In this paper, ARIMA models are used for forecasting international (outbound) tourism. Contrary to a causal oriented forecasting system, a time-series model accounts for patterns in the past movements of a variable and uses that information to predict its future movements. AR(I)MA time series models are designed for stationary time series. A stationary time series is one whose basic statistical properties such as the mean or variance remain constant over time. For the purpose of forecasting, the analyst must first identify the appropriate ARIMA model and transform and/or difference the data to produce stationarity. Achieving stationarity is reduced to the rather mechanical task of transforming a time series until the autocorrelations drop to zero with one, two, or three time lags.

Selected models were applied on yearly tourism import data of selected countries for the period 1975 to 1992. To achieve stationarity of the time series, only the absolute deviations from a quadratic trend are considered. The estimated model parameters fulfill the necessary stationarity and/or invertibility conditions.

Forecasts are completed until 1995. Based on the % changes of the original import data, an analysis of the results shows a recovery for the outbound travel of Western Europe in 1994 and 1995, following the weak year of 1993. From the selected

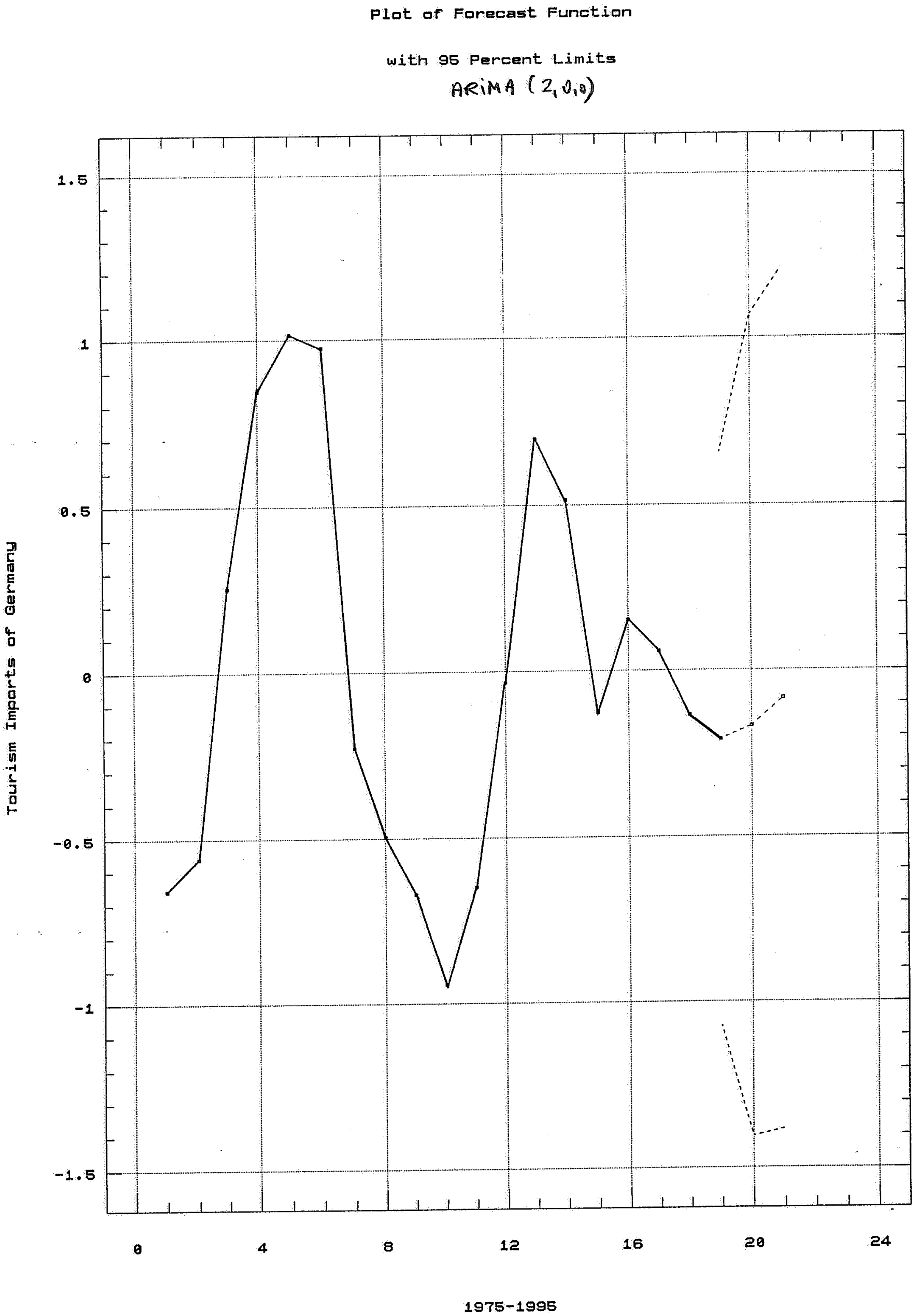
European countries, the growth expectations for Western Germany and the Netherlands are very poor. In all the other European countries such as the UK, France and Austria, the recovery will be to a greater or lesser extent strong. From the oversea countries, the tourism imports of Japan will expand very strongly: the growth rates for the tourism imports of Australia will be more moderate by far. The Canadian expenditures abroad seem to expand more slowly after 1993, falling in the years 1994 and 1995 below the trend.

Table 1

Tourism Imports 1993-1995

(Constant prices and exchange rates)

		1993	1994	1995
			%-changes	
Western Germany:	ARIMA (2,0,0)	0.2	1.5	1.8
	ARIMA (1,0,1)	0.8	0.9	1.1
United Kingdom:	ARIMA (2,0,0)	3.6	9.8	4.9
Austria:	ARIMA (0,0,1)	3.1	7.9	5.3
France:	ARIMA (0,0,1)	6.5	7.3	4.8
Netherlands	ARIMA (0,0,2)	0.2	-1.7	-0.4
Western Europe	ARIMA (2,0,0)	0.0	5.1	7.2
	ARIMA (1,0,1)	1.6	5.1	4.6
Canada	ARIMA (0,0,2)	10.0	2.2	6.4
Australia	ARIMA (0,0,1)	3.5	5.4	3.7
Japan	ARIMA (2,0,0)	12.1	22.4	12.5

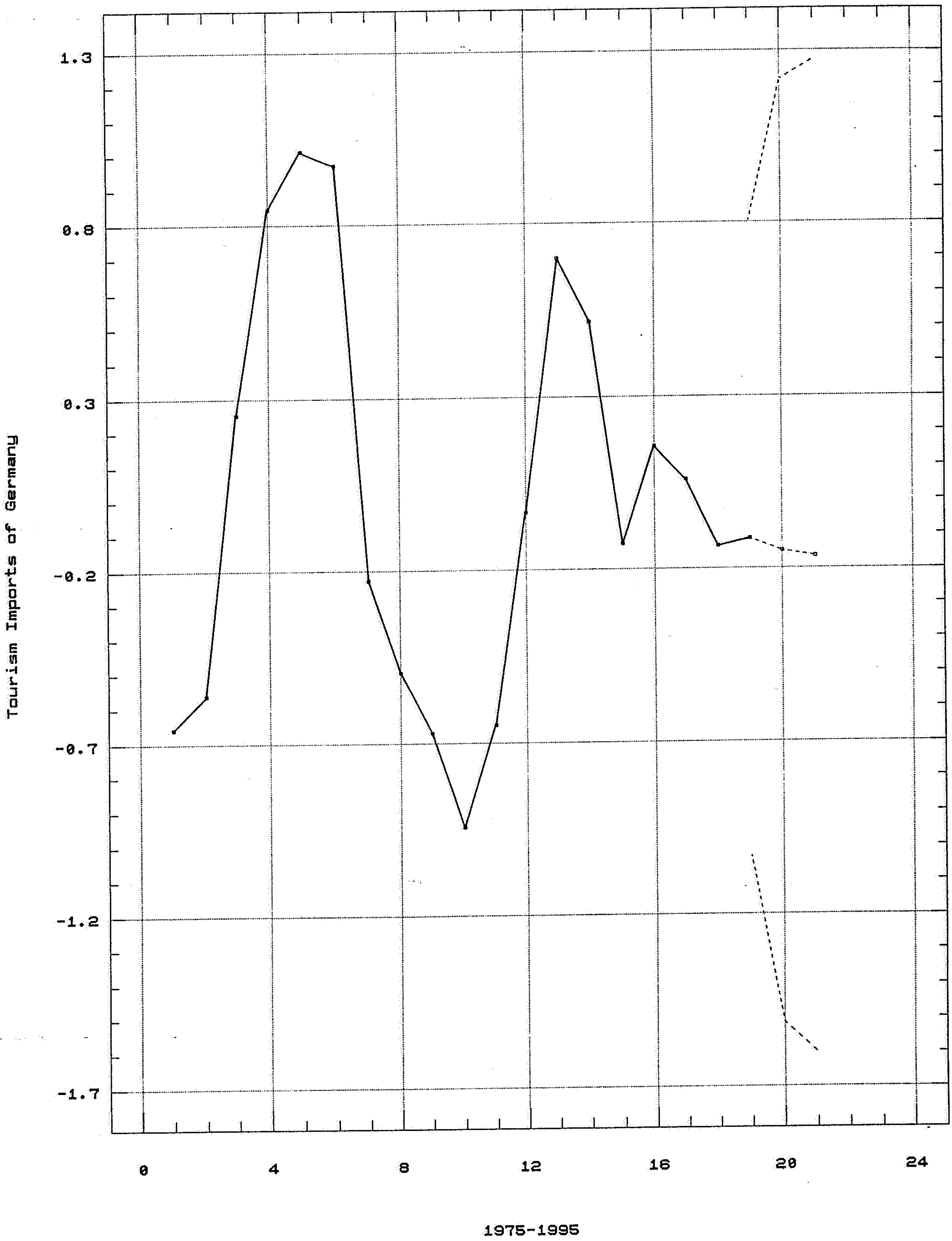


Trend deviations; deflated figures

Plot of Forecast Function

with 95 Percent Limits

ARIMA (1,0,1)

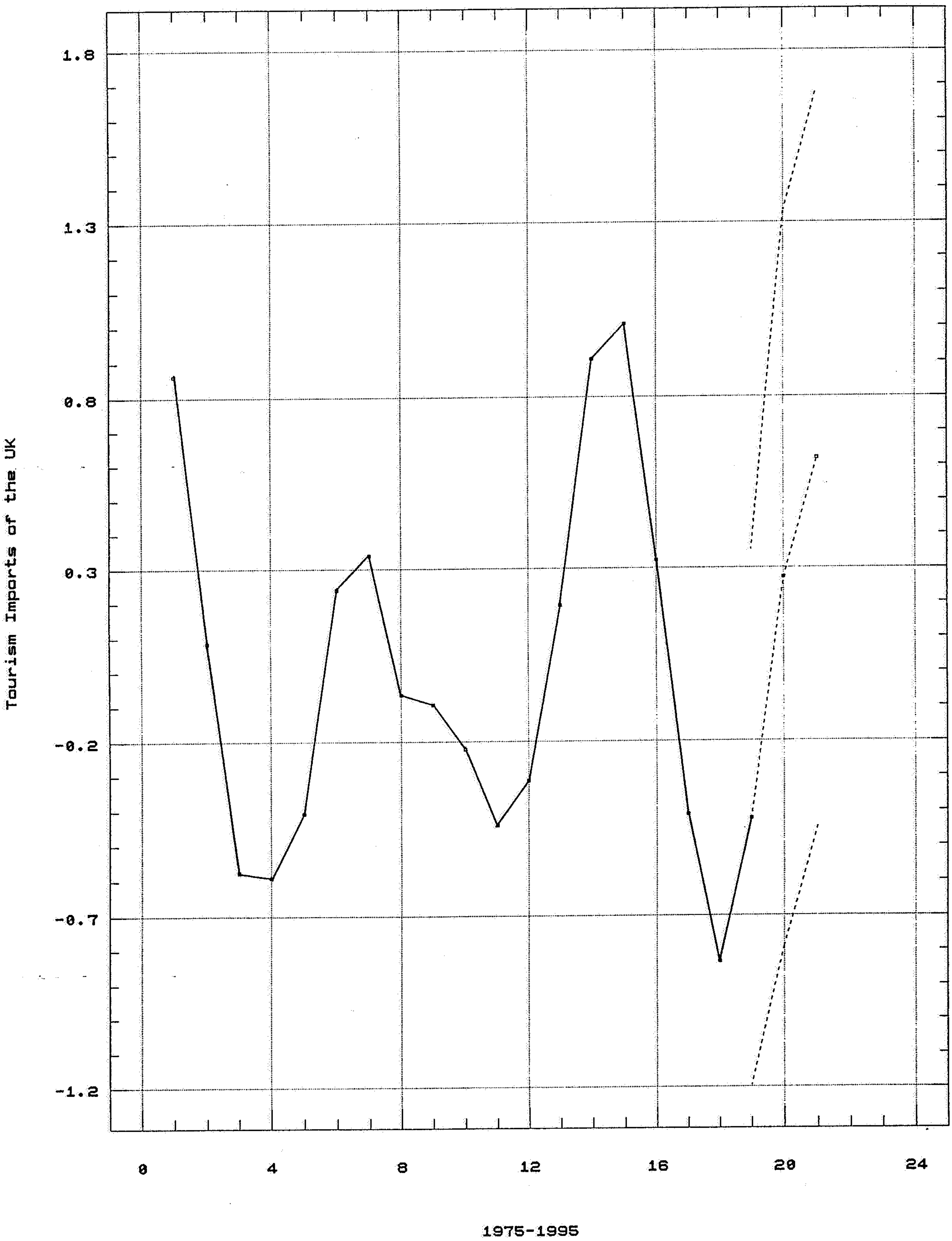


Trend deviations; deflated figures

Plot of Forecast Function

with 95 Percent Limits

ARIMA (2,0,0)

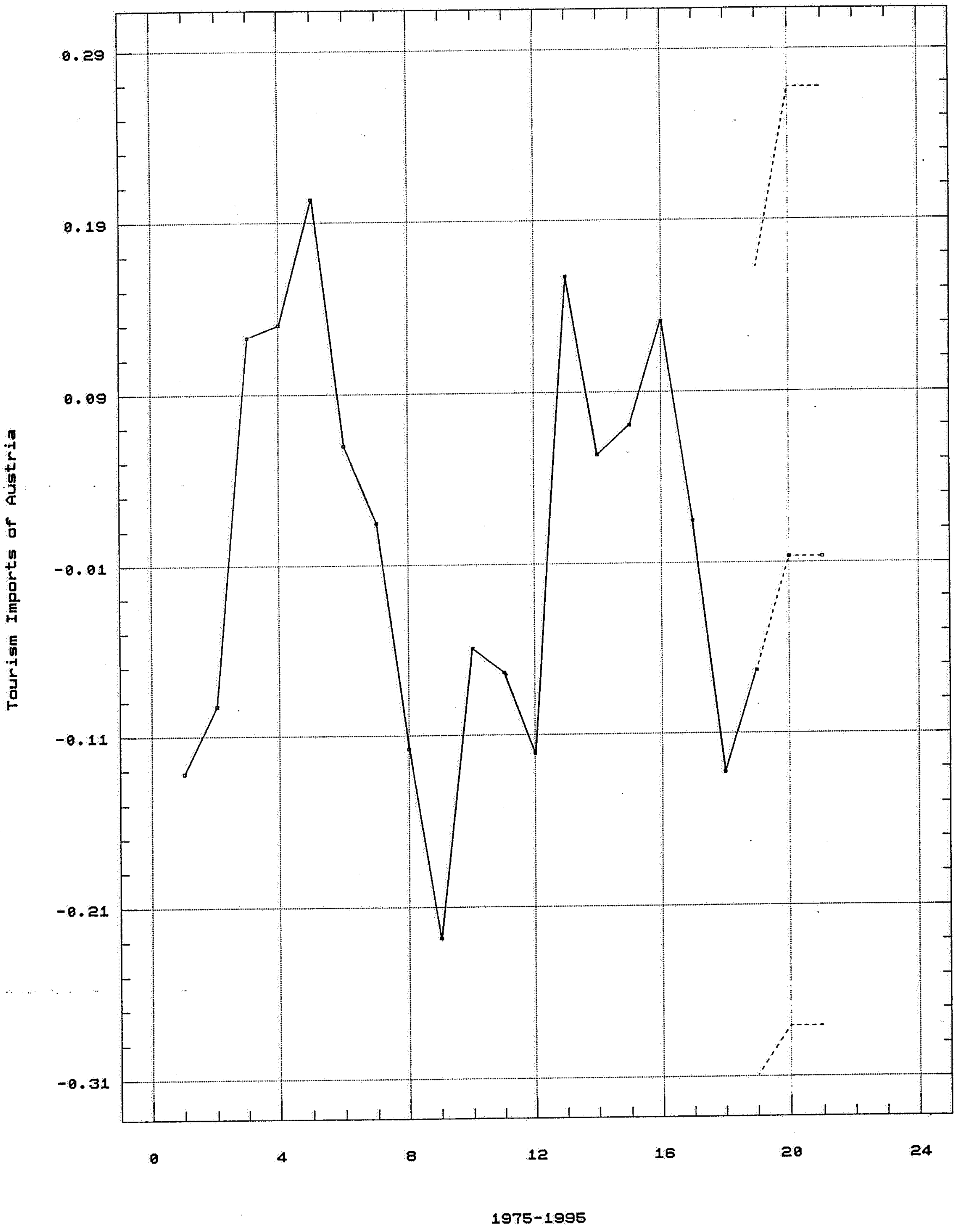


Trend deviations; deflated figures

Plot of Forecast Function

with 95 Percent Limits

ARIMA (0,0,1)

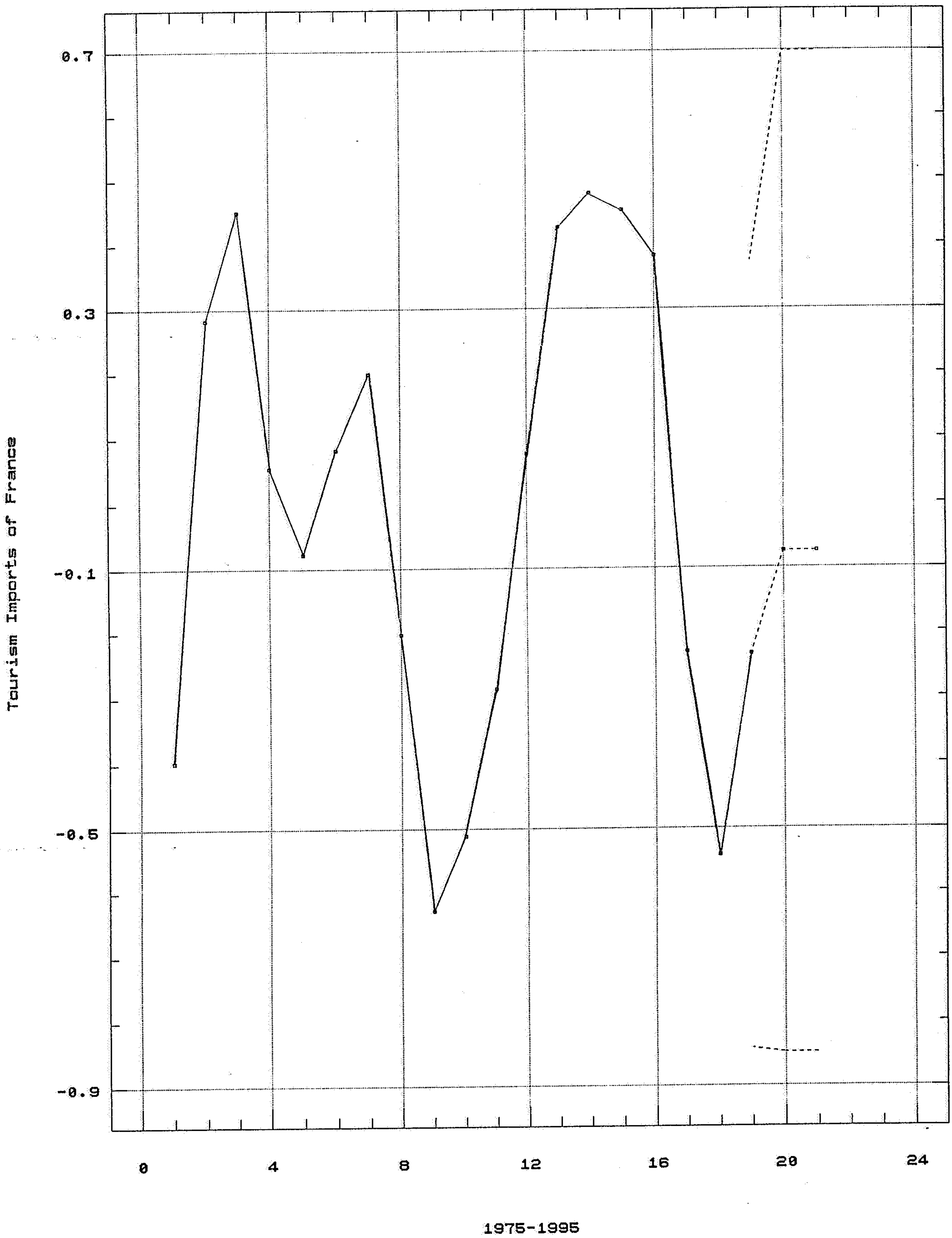


Trend deviations; deflated figures

Plot of Forecast Function

with 95 Percent Limits

ARIMA (0,0,1)

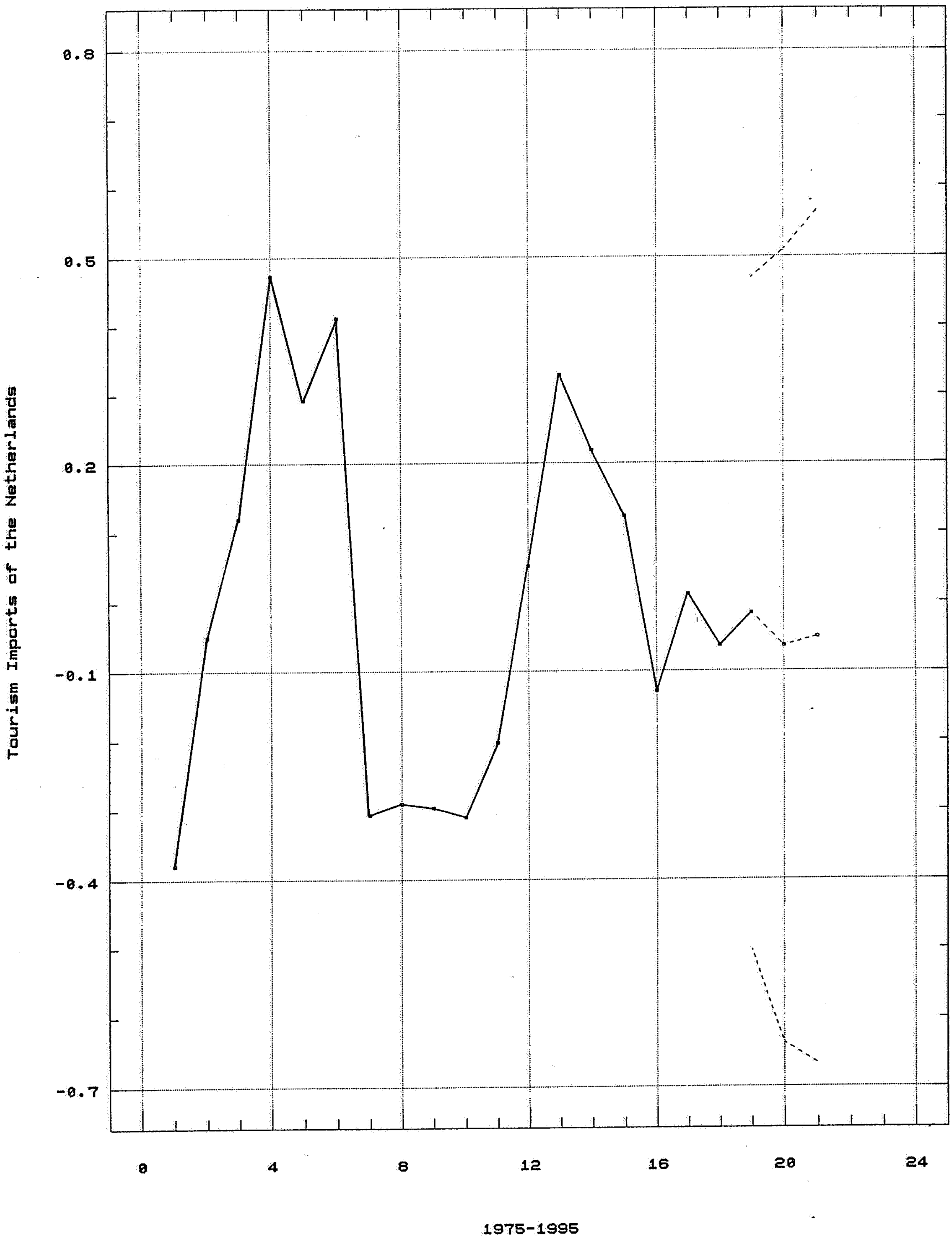


Trend deviations; deflated figures

Plot of Forecast Function

with 95 Percent Limits

ARIMA (0,0,2)

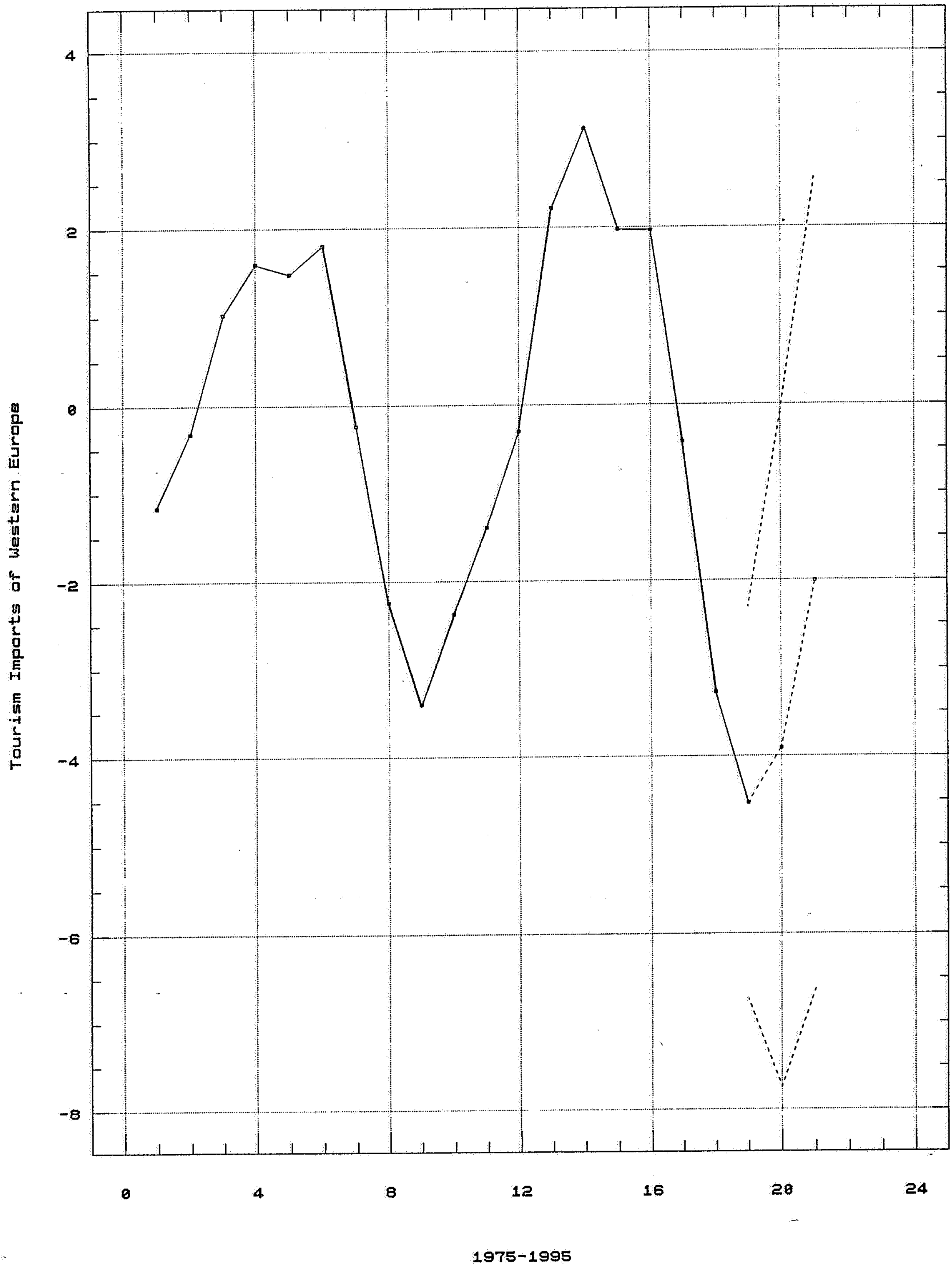


Trend deviations; deflated figures

Plot of Forecast Function

with 95 Percent Limits

ARIMA(2,0,0)

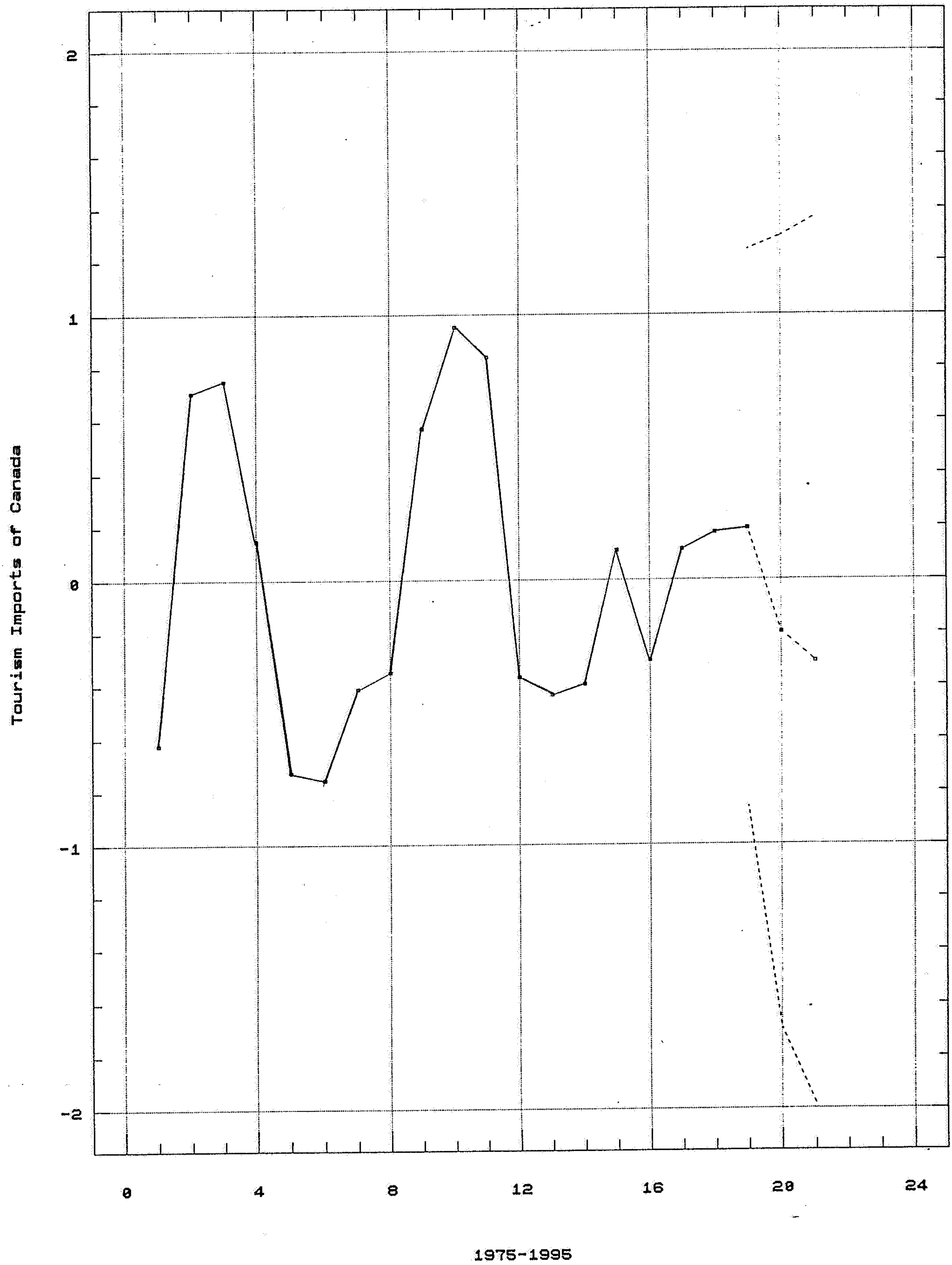


Trend deviations; deflated figures

Plot of Forecast Function

with 95 Percent Limits

ARIMA (0,0,2)

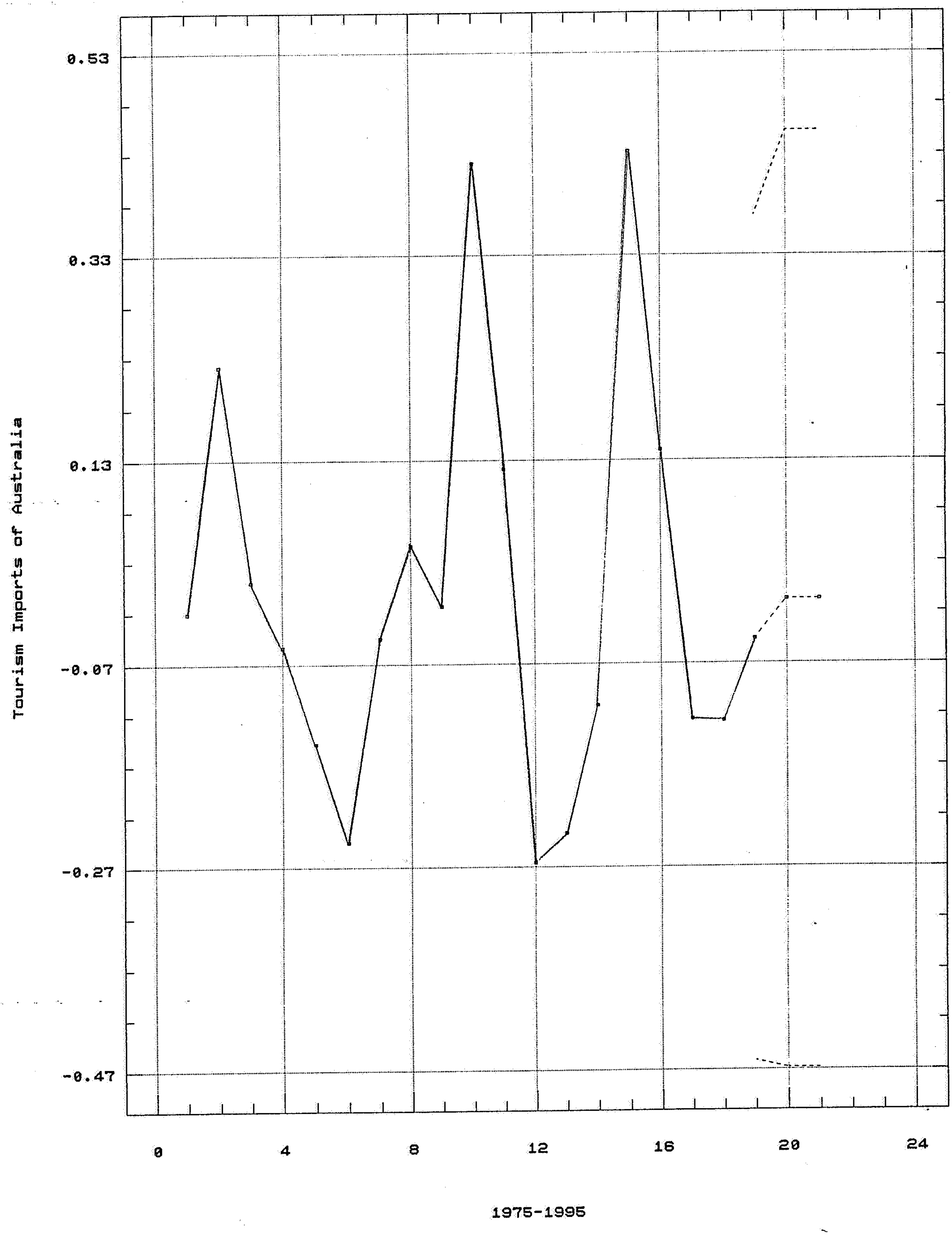


Trend deviations; deflated figures

Plot of Forecast Function

with 95 Percent Limits

ARIMA (0,0,1)

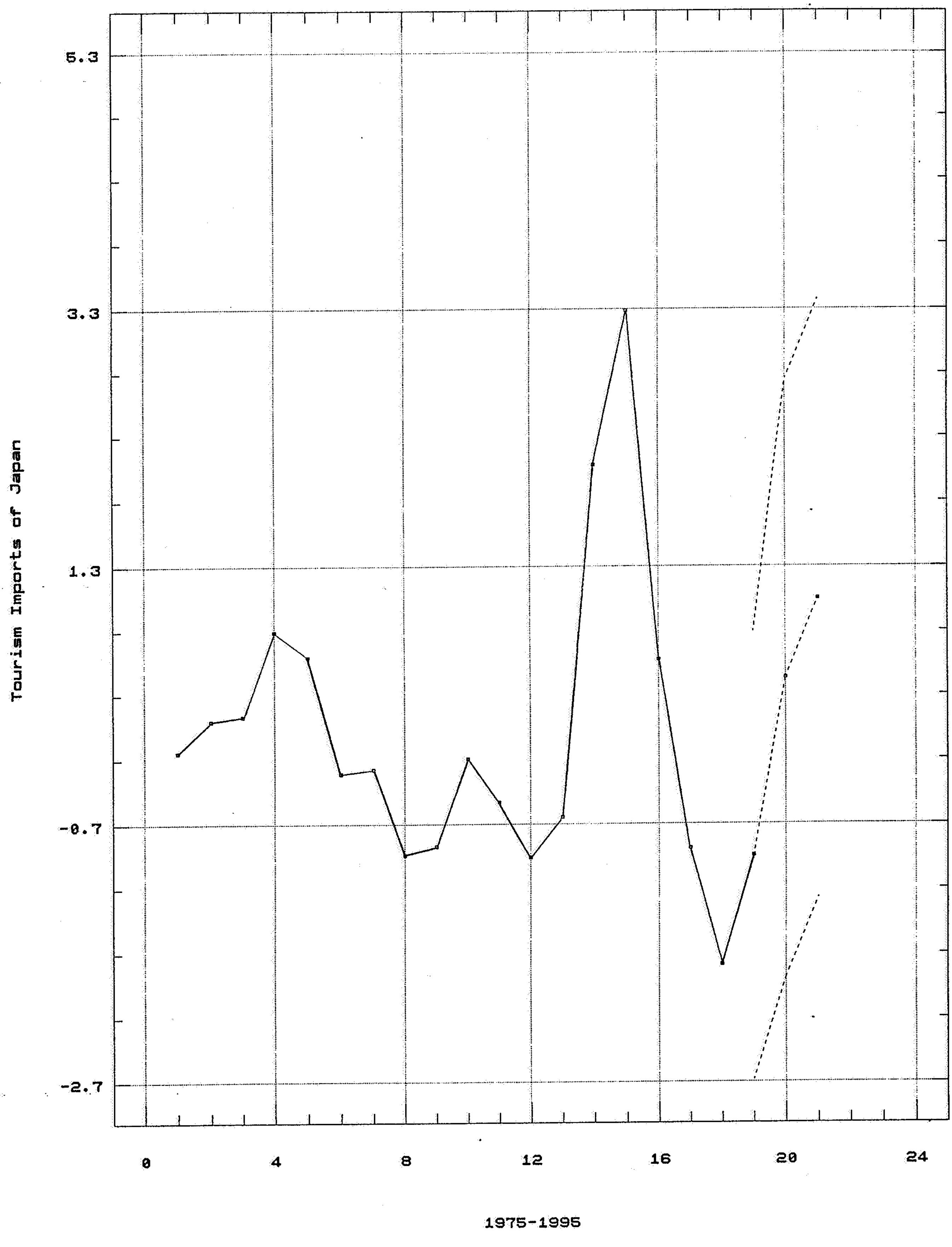


Trend deviations; deflated figures

Plot of Forecast Function

with 95 Percent Limits

ARIMA (2,0,0)



Trend deviations; deflated figures

ANNEX

Export models for Austria (1975-1992):

ARIMA (2,0,0); SEE = 0.158594

$$TD_t = -0.03565 + 1.38031*TD_{t-1} - 0.78230*TD_{t-2}$$

ARIMA (2,0,0); outlier corrected; SEE = 0.106874

$$TD_t = -0.01956 + 1.41098*TD_{t-1} - 0.83106*TD_{t-2}$$

ARIMA (1,0,1); SEE = 0.192825

$$TD_t = -0.04586 + 0.68601*TD_{t-1} + e_t + 0.56610*e_{t-1}$$

Tourism Exports 1993-1995

(Constant prices and exchange rates)

		1993	1994	1995
		%-changes		
Austria:	ARIMA (2,0,0)	-1.5	5.7	6.3
	outlier corr.	-2.8	2.2	4.5
	ARIMA (1,0,1)	-2.2	5.3	5.3

References

Box, G., E., P., Jenkins, G.,M., Time Series Analysis: Forecasting and Control, San Francisco, 1970.

Fritz, R.,G., et al. Combining Time Series and Econometric Forecast of Tourism Activity, Annals of Tourism Research, 1984, 219-229.

Geurts, M.,D., Ibrahim, I., B., Comparing the Box-Jenkins Approach with the Exponentially Smoothed Forecasting Model: Application to Hawaii Tourists, Journal of Marketing Research, 1975, 182-188.

Jarrett, J., Business Forecasting Methods, second Edition, Cambridge, Mass., 1991.

Makridakis, S., et al., Forecasting: Methods and Applications, New York, 1983.

Pack, J., A Practical Overview of ARIMA Models for Times Series Forecasting in: Makridakis, S., et al., The Handbook of Forecasting, A Manager's Guide, second Edition, New York, 1987.

Pindyck, R.S., Rubinfeld, D., L., Econometric Models and Economic Forecasts.

Witt, St., Martin, C.,A., Demand Forecasting in Tourism and Recreation, in: Cooper, C.,P., Progress in Tourism, Recreation and Hospitality Management, London, 1987.1

Witt, St. et al.,The Management of International Tourism, London, 1991.

Witt, St., Witt, Ch.,A., Modeling and Forecasting Demand in Tourism, London, 1992.

Eigentümer, Herausgeber und Verleger: Gemeinnütziger Verein "Österreichisches Institut für Wirtschaftsforschung", Wien 3, Arsenal, Objekt 20. Postanschrift: A-1103 Wien, Postfach 91. Tel. (1) 798 26 01-0, Fax (1) 798 93 86. Vorstand: Präsident: Ing. Leopold Maderthaner, Vizepräsidenten: Eleonora Hostasch, Univ.-Prof. Dr. Erich Streißler, Geschäftsführer: Prof. Dr. Helmut Kramer.

Satz und Druck: Österreichisches Institut für Wirtschaftsforschung.

Verkaufspreis: S 100,–.