Statistics, O.S.

Netherlands School of Economics

ECONOMETRIC INSTITUTE

Report 6507

THE INFORMATION APPROACH TO THE PREDICTION OF INTERREGIONAL TRADE FLOWS

by Pedro Uribe, C.G. de Leeuw and H. Theil

June 3, 1965 Preliminary and confidential

THE INFORMATION APPROACH
TO THE PREDICTION OF INTERREGIONAL TRADE FLOWS

by Pedro Uribe,[1] C.G. de Leeuw and H. Theil

Contents

1. INTRODUCTION

In a recent article [5] published in the Cahiers économiques de Bruxelles Professor J. Waelbroeck analyzed world trade flows in a very interesting way. He took total exports to and total imports from each region as given and applied the RAS method - originally constructed for input-output analysis - to analyze the extent to which flows known from an earlier table can account for flows in a later year. The approach has some resemblance to certain aspects of bivariate information theory. The present paper takes this theory as a starting point to provide an alternative method of analysis. The empirical part of the paper is based on the same data as those which are used by Waelbroeck. It turns out that the two methods yield virtually equivalent results, which is not really surprising, but which sheds an interesting light on the nature of the RAS method.

2. INTERNATIONAL TRANSACTION SHARES FROM THE STANDPOINT OF INFORMATION THEORY

We divide the world into n regions and indicate by X_{ij} the flow of exports from region i to region j, by $X_{i.} = \Sigma X_{ij}$ (sum over j) total exports from region i, by $X_{.j} = \Sigma X_{ij}$ (sum over i) total imports by region j, and by $X_{..} = \Sigma \Sigma X_{ij}$ world trade. We divide all these flows

[1] Mr. Uribe wrote this article while being a guest of the Econometric Institute under a grant of the International Technical Assistance of the Dutch government.

by world trade, so that we obtain

(2.1) $$x_{ij} = \frac{X_{ij}}{X_{..}} \qquad x_{i.} = \frac{X_{i.}}{X_{..}} \qquad x_{.j} = \frac{X_{.j}}{X_{..}}$$

which are, respectively, the share of the flow from i to j in world trade, the share of i in world exports, and the share of j in world imports. Since the x_{ij} are nonnegative and add up to 1, they can be regarded as probabilities of a bivariate distribution, and the $x_{i.}$ and $x_{.j}$ as the corresponding marginal probabilities.

We assume that the flows X_{ij} refer to some previous year. Let it be our task to predict the flows for some later year, to be denoted by Y_{ij}, under the condition that the marginal totals $Y_{i.}$, $Y_{.j}$ are given. Hence it is sufficient to predict the shares $y_{ij} = Y_{ij}/Y_{..}$, given all $y_{i.} = Y_{i.}/Y_{..}$ and $y_{.j} = Y_{.j}/Y_{..}$, where $Y_{..}$ is the same function of Y_{ij} as $X_{..}$ is of X_{ij}.

One prediction method is

(2.2) $$\hat{y}^1_{ij} = y_{i.}y_{.j} \qquad i, j = 1, \ldots, n$$

which amounts to the assumption of import-export independence. It means that exports from i to j are supposed to be considerable when i exports much and j imports much, and that the flow from i to j becomes less when either i exports little or j imports little or both. Although there is undoubtedly some truth in this assumption, it is clearly of a very approximate nature. Some region i may export little to some other region j, in spite of substantial values of $y_{i.}$ and $y_{.j}$, simply because their distance is sizable or because their governments have troubles with each other. Now some of these causes have a quasi-permanent character, such as distance and possibly political problems. It is therefore conceivable that we can improve on the prediction method (2.2) by taking account of the flow distribution in an earlier year. So we return to the year of our x_{ij}.

For that year it will, generally, not be true that $x_{ij} = x_{i.}x_{.j}$ holds for all pairs (i, j). The left-hand side will exceed the right-hand side for some pairs - those whose trade is above the independence level; the left-hand side will be lower than the right-hand side for other pairs - those who trade with each other at below-independence level. Consider

(2.3) $$\log \frac{x_{ij}}{x_{i.}x_{.j}}$$

which is known in information theory (see [2], pp. 30-31) as the <u>mutual information</u> between i and j; that is, in this case, between the exporting region i and the importing region j. It will be positive when the trade from i to j is above the independence level, negative when it is below that level. An obvious prediction method, presumably superior to (2.2), amounts to assuming that this mutual information remains unchanged. That is,

$$\frac{y'_{ij}}{y_{i.}y_{.j}} = \frac{x_{ij}}{x_{i.}x_{.j}} \qquad i, j = 1, \ldots, n$$

where y'_{ij} is the forecast of y_{ij}. We thus have:

$$(2.4) \qquad y'_{ij} = \frac{y_{i.}y_{.j}}{x_{i.}x_{.j}} x_{ij} \qquad i, j = 1, \ldots, n$$

Note, however, that the y'_{ij} do not necessarily add up to 1. This can be remedied easily by adjusting them proportionally:

$$(2.5) \qquad \hat{y}^2_{ij} = \frac{y'_{ij}}{\sum_{h=1}^{n} \sum_{k=1}^{n} y'_{hk}} \qquad i, j = 1, \ldots, n$$

An important question is: Now that we have obtained two competing sets of forecasts, how can we evaluate their merits when they are both compared with the observed y_{ij} values? We can, of course, compare all n^2 pairs of individual prediction errors, but this is quite inconvenient unless n is very small. Moreover, the individual comparisons will usually not lead to a straightforward overall evaluation of all n^2 predictions combined. An alternative method, also derived from information theory, seems preferable. We regard the forecast $\hat{y}_{ij}$ as a "prior" probability, which is followed later on by a "message" on the actual values, i.e., the observed values y_{ij}, which are regarded as "posterior" probabilities. We consider the forecasts to be accurate when the realizations (the posterior probabilities) give little information, given the forecasts made (the prior probabilities). Conversely, we consider the forecasts to be inaccurate when the realizations contain a great deal of information, given the forecasts. The information content of a message leading to the posterior probabilities y_{ij}, given the prior probabilities $\hat{y}_{ij}$, is defined in information theory as

$$(2.6) \qquad I = \sum_{i=1}^{n} \sum_{j=1}^{n} y_{ij} \log \frac{y_{ij}}{\hat{y}_{ij}}$$

This measure, which takes larger values when the differences between corresponding values $\hat{y}_{ij}$ and y_{ij} become larger, is known as the <u>information inaccuracy</u> of the forecasts $\hat{y}_{ij}$ with respect to the realizations y_{ij}; see [4]. We shall use 2 as the base of our logarithms, as is usual in information theory, in which case I is expressed in "bits." Note that the independence prediction method (2.2) leads to the following value for I:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} y_{ij} \log \frac{y_{ij}}{y_{i.} y_{.j}}$$

On comparing this with the discussion below (2.3) we conclude that this is simply the expected mutual information of the n^2 flows y_{ij} in the year to be predicted.

3. SOME EMPIRICAL RESULTS; A THIRD PREDICTION METHOD

We shall now apply the prediction procedures (2.2) and (2.5) to import and export data of the years 1938, 1948, 1951-52, and 1959-60 of the following 8 regions:[2]

- North America
- Latin America
- Germany
- Other E.E.C. countries
- United Kingdom
- Other E.F.T.A. countries
- Communist countries
- Rest of the world

For each prediction method we can take any of the four years as a base and predict (forecast or "backcast") any of the other years. For the simple independence method (2.2) there is, of course, only one set of predictions for each year. The following are its inaccuracy values:

1938	.2061 bit
1948	.2540 bit
1951-52	.3479 bit
1959-60	.3532 bit

The values are increasing over time, which implies a trend away from independence.

The figures just given must be regarded as a kind of upper limit to the inaccuracy value obtained by the use of more sophisticated prediction procedures. Table 1 shows the extent to which this is indeed the case. Each cell of the table consists of three figures, the first of which refers to $\hat{y}^2_{ij}$ as defined in (2.5); the second will be discussed at the end of this section, the third in the next section. On

---

2 The data are the same as those used by M. Waelbroeck except that Japan, which was taken by him as a separate region, is allocated here to the rest of the world. This is motivated by the zero flow from Other E.E.C. countries to Japan in 1938, followed by positive flows in later years, which would have led to infinite inaccuracy values. Note further that Germany should be interpreted as Western Germany after the War, and that Yugoslavia has been allocated to the rest of the world rather than to the Communist countries.

comparing the $\hat{y}^2_{ij}$ figures with those of $\hat{y}_{ij}$ discussed above, we find that they are considerably lower except when 1938 is backcast by means of the early postwar data. Hence it is indeed true that we gain in prediction accuracy when we take account of the flow pattern of an earlier year.

TABLE 1. INFORMATION INACCURACY VALUES OF FOUR ALTERNATIVE PREDICTION METHODS

Base year	Year to be predicted			
	1938	1948	1951-52	1959-60
1938		1146 995 996	1616 1553 1552	1728 1701 1702
1948	2908 2702 2617		964 906 900	1747 1479 1439
1951-52	2478 2402 2367	994 948 948		456 338 340
1959-60	1572 1511 1492	920 728 735	320 247 246	

Note. All values are in 10^{-4} bits.

The prediction method (2.5) has been derived from y'_{ij} as defined in (2.4) in such a way that it satisfies the constraint that the double sum over $\hat{y}^2_{ij}$ be equal to 1. It is important to realize, however, that the single sum of these $\hat{y}^2_{ij}$ over j is not equal to the $y_{i.}$ which is taken as given, and that the sum of $\hat{y}^2_{ij}$ over i is not equal to $y_{.j}$. This is pursued in Table 2, which contains $y'_{i.} = \Sigma\, y'_{ij}$ (sum over j) and $y'_{.j} = \Sigma\, y'_{ij}$ (sum over i) as well as the observed $y_{i.}$, $y_{.j}$. It turns out that the double sum $y'_{..}$ is close to 1 in all cases, so that the correction (2.5) is of minor importance (at most 2 per cent and frequently much less). However, the discrepancies $y'_{i.} - y_{i.}$ and $y'_{.j} - y_{.j}$ are larger. This difference between the behavior of subtotals and grand total is not really surprising, which can be argued as follows.

Let us write

$$(3.1) \qquad y_{i.} = (1 + \delta_i)x_{i.} \qquad\qquad y_{.j} = (1 + \varepsilon_j)x_{.j}$$

Then we find after summation over i and over j, respectively

$$(3.2) \qquad \sum_{i=1}^{n} x_{i.}\delta_i = \sum_{j=1}^{n} x_{.j}\varepsilon_j = 0$$

Going back to (2.4), we find for the unadjusted forecast y'_{ij}:

$$y'_{ij} = (1 + \delta_i)(1 + \varepsilon_j)x_{ij}$$

so that the marginal sums are

$$y'_{i.} = (1 + \delta_i) \sum_{j=1}^{n} (1 + \varepsilon_j)x_{ij} = (1 + \delta_i)x_{i.} + (1 + \delta_i) \sum_{j=1}^{n} x_{ij}\varepsilon_j$$

$$= y_{i.} + (1 + \delta_i) \sum_{j=1}^{n} x_{ij}\varepsilon_j$$

and

$$y'_{.j} = (1 + \varepsilon_j) \sum_{i=1}^{n} (1 + \delta_i)x_{ij} = y_{.j} + (1 + \varepsilon_j) \sum_{i=1}^{n} x_{ij}\delta_i$$

This shows that the discrepancies $y'_{i.} - y_{i.}$ and $y'_{.j} - y_{.j}$ are of the same order of magnitude as the ε's and the δ's, respectively. Consider now the double sum:

$$(3.3) \qquad y'_{..} = \sum_{i=1}^{n} y'_{i.} = \sum_{i=1}^{n} y_{i.} + \sum_{i=1}^{n} (1 + \delta_i) \sum_{j=1}^{n} x_{ij}\varepsilon_j$$

$$= 1 + \sum_{i=1}^{n} \sum_{j=1}^{n} x_{ij}\varepsilon_j + \sum_{i=1}^{n} \sum_{j=1}^{n} x_{ij}\delta_i\varepsilon_j$$

$$= 1 + \sum_{i=1}^{n} \sum_{j=1}^{n} x_{ij}\delta_i\varepsilon_j$$

where use is made of

$$\sum_{i=1}^{n} \sum_{j=1}^{n} x_{ij}\varepsilon_j = \sum_{j=1}^{n} x_{.j}\varepsilon_j = 0$$

see (3.2). The result (3.3) shows that $y'_{..}$ is equal to 1 apart from a discrepancy which is of the second order of smallness in the δ's and ε's.

Given that the forecast (2.5) does not satisfy its marginal constraints, one may wonder whether we can adjust the y'_{ij} defined in (2.4) in such a way that it satisfies both the "total" constraint [as 2.5 already does] and also each of the marginal constraints. Since our criterion for the merits of forecasts is the information criterion (2.6), an obvious procedure is to define a third set of forecasts $\hat{y}^3_{ij}$ which minimize the information inaccuracy with respect to $\hat{y}^2_{ij}$:

$$(3.4) \qquad \sum_{i=1}^{n} \sum_{j=1}^{n} \hat{y}^2_{ij} \log \frac{\hat{y}^3_{ij}}{\hat{y}^2_{ij}}$$

subject to

TABLE 2. OBSERVED AND PREDICTED ROW AND COLUMN TOTALS

$(y_{i.}, y'_{i.}, y_{.j}, y'_{.j} \times 10^4)$

Year predicted	Base year of prediction	North America	Latin America	Germany	Other E.E.C. countries	United Kingdom	Other E.F.T.A. countries	Communist countries	Rest of the world	Total
		Row totals (total regional exports divided by world trade)								
1938	observed (1938)	1674	729	962	897	1032	742	1015	2949	10000
	1948	1701	672	888	915	1000	768	1051	2977	9972
	1951-52	1543	660	907	962	949	860	982	3177	10040
	1959-60	1669	703	922	933	992	824	793	3239	10076
1948	observed (1948)	2731	1136	123	1011	1099	569	644	2686	10000
	1938	2961	1132	138	962	1192	538	604	2511	10039
	1951-52	2774	1130	132	1017	1096	573	573	2703	9998
	1959-60	3043	1160	134	968	1141	552	465	2758	10222
1951-52	observed (1951-52)	2355	918	464	1238	888	609	824	2704	10000
	1938	2501	933	502	1169	982	557	782	2576	10003
	1948	2372	913	422	1223	905	597	888	2653	9974
	1959-60	2571	945	478	1184	931	586	659	2770	10123
1959-60	observed (1959-60)	2012	696	877	1394	797	645	1205	2374	10000
	1938	2016	704	976	1329	854	615	1224	2267	9984
	1948	2020	664	810	1382	780	664	1587	2252	10160
	1951-52	1858	673	856	1398	749	680	1580	2358	10152

		Column totals (total regional imports divided by world trade)								
1938	observed (1938)	1198	588	1036	1224	1655	661	714	2923	10000
	1948	988	437	790	1518	1751	824	889	2776	9972
	1951-52	1065	524	971	1316	1668	756	846	2895	10040
	1959-60	1211	554	1008	1225	1749	648	662	3020	10076
1948	observed (1948)	1653	965	276	1469	1243	729	627	3038	10000
	1938	2012	1050	287	1258	1333	593	511	2994	10039
	1951-52	1822	1024	277	1370	1227	646	529	3103	9998
	1959-60	2027	1095	284	1294	1334	587	434	3166	10222
1951-52	observed (1951-52)	1728	839	454	1292	1098	665	771	3154	10000
	1938	1919	892	473	1203	1158	602	680	3075	10003
	1948	1581	761	420	1434	1122	748	860	3047	9974
	1959-60	1938	892	470	1223	1171	594	620	3215	10123
1959-60	observed (1959-60)	1667	635	768	1314	891	743	1193	2789	10000
	1938	1667	678	792	1334	895	780	1151	2688	9984
	1948	1367	520	656	1595	887	955	1583	2597	10160
	1951-52	1470	602	736	1374	850	852	1530	2738	10152

Note: All figures are to be multiplied by 10^{-4}

$$\sum_{j=1}^{n} \hat{y}^3_{ij} = y_{i.} \qquad i = 1, \ldots, n$$

(3.5)

$$\sum_{i=1}^{n} \hat{y}^3_{ij} = y_{.j} \qquad j = 1, \ldots, n$$

This is what may be called a "two-stage information forecast." The first stage consists of computing preliminary forecasts y'_{ij} on the basis of constant mutual information values. These preliminary forecasts are adjusted proportionally so as to obtain $\hat{y}^2_{ij}$, after which $\hat{y}^3_{ij}$ is obtained by minimizing, subject to (3.5), the information inaccuracy (3.4) in which $\hat{y}^2_{ij}$ is the "actual" and $\hat{y}^3_{ij}$ the "predicted" value.

The actual task of minimizing (3.4) subject to (3.5) is rather awkward. The differences between the various forecasts are rather small, however, so that we may decide to expand the inaccuracy according to powers of $\hat{y}^3_{ij} - \hat{y}^2_{ij}$. The leading nonzero term is quadratic and is proportional to a chi-square:

(3.6)
$$\sum_{i=1}^{n} \sum_{j=1}^{n} \frac{(\hat{y}^3_{ij} - \hat{y}^2_{ij})^2}{\hat{y}^2_{ij}}$$

So we minimize the quadratic form (3.6) subject to the linear constraints (3.5), which is a straightforward procedure.[3] The information inaccuracy values are shown in each call of Table 1 (second number). The results indicate that there is a uniform improvement over the method (2.5). We conclude that it is important to take the marginal constraints seriously.

4. THE FOURTH METHOD: RAS

The method employed by M. Waelbroeck himself is the RAS method, which was originally designed by Stone and Brown [3] for the adjustment of input coefficient matrices. It amounts to the computation of n values r_i and n values s_j such that

(4.1)
$$\hat{y}^4_{ij} = r_i x_{ij} s_j \qquad i, j = 1, \ldots, n$$

satisfies the constraints

[3] Stephan's method has been used; see e.g. Deming [1], pp. 121-124. Note that it is immaterial for $\hat{y}^3_{ij}$ whether we use $\hat{y}^2_{ij}$ in (3.6) or y'_{ij}. Hence the proportional adjustment of the preliminary forecasts is not really necessary.

$$(4.2)\qquad \sum_{j=1}^{n} \hat{y}^4_{ij} = r_i \sum_{j=1}^{n} x_{ij}s_j = y_{i.} \qquad i = 1, \ldots, n$$

$$\sum_{i=1}^{n} \hat{y}^4_{ij} = s_j \sum_{i=1}^{n} x_{ij}r_i = y_{.j} \qquad j = 1, \ldots, n$$

The inaccuracy values are shown on the third line of each cell of Table 1. They turn out to be of the same order of magnitude as those of the first line, which is not really surprising. If we would identify r_i with $y_{i.}/x_{i.}$ and s_j with $y_{.j}/x_{.j}$, the forecast (4.1) would be the same as y'_{ij} defined in (2.4).[4] This identification is not really correct, of course, since $\hat{y}^4_{ij}$ does and y'_{ij} does not satisfy the marginal constraints. But if we adjust y'_{ij} such that these constraints are satisfied, it stands to reason that the result $\hat{y}^3_{ij}$ will not differ very much from $\hat{y}^4_{ij}$. Therefore, the RAS method can be regarded as being approximately equivalent to the information criterion when the y'_{ij} do not violate the marginal constraints too seriously.

5. THE INFORMATION INACCURACY OF FORECASTS OF IMPORT AND EXPORT DISTRIBUTIONS FOR INDIVIDUAL REGIONS

Until now we considered the problem of the prediction of all n^2 shares y_{ij} in world trade. We may also be interested in the prediction of the destination distribution of all exports from a given region i to all regions, $y_{ij}/y_{i.}$, $j = 1, \ldots, n$, or in the origin distribution of all imports to a given region j, $y_{ij}/y_{.j}$, $i = 1, \ldots, n$. The corresponding inaccuracy values are

$$(5.1)\qquad I_{i.} = \sum_{j=1}^{n} \frac{y_{ij}}{y_{i.}} \log \frac{y_{ij}/y_{i.}}{\hat{y}_{ij}/\hat{y}_{i.}} = \frac{1}{y_{i.}} \sum_{j=1}^{n} y_{ij} \log \frac{y_{ij}}{\hat{y}_{ij}} - \log \frac{y_{i.}}{\hat{y}_{i.}}$$

in the case of exports, and in that of imports:

$$(5.2)\qquad I_{.j} = \sum_{i=1}^{n} \frac{y_{ij}}{y_{.j}} \log \frac{y_{ij}/y_{.j}}{\hat{y}_{ij}/\hat{y}_{.j}} = \frac{1}{y_{.j}} \sum_{i=1}^{n} y_{ij} \log \frac{y_{ij}}{\hat{y}_{ij}} - \log \frac{y_{.j}}{\hat{y}_{.j}}$$

4 Actually, the iterative procedure which leads to the RAS forecasts starts in the first round with precisely these r_i and s_j values.

When the forecasts satisfy the marginal constraints (the case of $\hat{y}^3_{ij}$ and $\hat{y}^4_{ij}$) the second term on the second line of both (5.1) and (5.2) will vanish.

The results are presented in Table 3 for all forecasts (not for the backcasts). By and large, they show the same pattern as that of the "total" inaccuracy. Each cell of the table contains four numbers. The first corresponds to the independence prediction (2.2), the second to $\hat{y}^2_{ij}$ as defined in (2.5). This second value is in almost all cases considerably less than the first. The third value of each cell corresponds to $\hat{y}^3_{ij}$, the fourth to RAS. They are both better than $\hat{y}^2_{ij}$ in a majority of all cases, but the majority is not overwhelming (less than 60 per cent). When comparing the various cells, we find that the largest figures are those of the Communist countries. This is the effect of a particular trade policy as will be explained in some detail in the next section.

6. TRACING THE E.E.C. AND THE COMMUNIST POLICY EFFECTS

In concluding we want to pay some attention to the possibility of using information analysis to the problem of tracing certain special developments. For this purpose we return to the mutual information concept (2.3),

$$\text{(6.1)} \qquad \log \frac{x_{ij}}{x_{i.}x_{.j}}$$

As stated above, this logarithm is positive when the flow from i to j is above the independence level, negative when it is below that level, and it takes algebraically larger values when the flow from i to j increases relative to the total exports of i and the total imports of j. It seems clear that the development of this logarithmic ratio should enable us, at least in principle, to draw interesting conclusions about the changes in the trade pattern. In the present case we have only four observations on each mutual information, but it is nevertheless instructive to consider a few cases.

One such case is that of the Communist countries (region 7). For the trade between these countries we consider (6.1) with i = j = 7, which takes the following values:

1938	.619
1948	2.663
1951-52	2.973
1959-60	2.408

These figures clearly indicate that the countries involved concentrated their trade among each other after the War. Of course, this feature is

TABLE 3. INFORMATION INACCURACIES FOR EXPORT SHARE AND IMPORT SHARE PREDICTIONS

Year predicted	Base year of prediction	North America	Latin America	Germany	Other E.E.C. countries	United Kingdom	Other E.F.T.A. countries	Communist countries	Rest of the world
		Export share information inaccuracies							
1948	1938	1567	2538	10533	1550	4550	1757	8000	1571
		954	442	7208	459	727	962	5165	744
		931	294	6754	263	595	960	4576	681
		928	293	6803	262	599	957	4580	682
1951-52	1938	2527	4478	5006	1571	4549	2250	14575	1123
		1180	429	2095	283	860	876	10388	567
		1243	442	2241	240	752	996	9575	627
		1241	446	2219	239	760	1004	9559	630
	1948	2527	4478	5006	1571	4549	2250	14575	1123
		1054	416	6330	271	317	720	1076	673
		972	279	5714	238	444	565	1295	653
		976	280	5599	239	436	560	1291	652
1959-60	1938	2814	4153	3880	1751	3462	2248	11916	994
		1117	153	2234	241	432	1485	8462	404
		1125	140	2365	297	483	1501	7986	499
		1126	140	2352	301	478	1502	7991	503
	1948	2814	4153	3880	1751	3462	2248	11916	994
		1776	462	6809	740	548	1371	1071	1326
		1818	215	5539	897	772	863	1185	957
		1799	212	5233	905	757	843	1160	940
	1951-52	2814	4153	3880	1751	3462	2248	11916	994
		378	402	482	640	360	160	483	174
		397	449	342	738	296	132	331	92
		400	449	339	743	297	128	345	90

		Import share information inaccuracies							
1948	1938	2403	4604	4486	1058	3172	2551	8121	1086
		297	690	9350	768	1156	533	4343	271
		210	691	10126	719	926	480	4237	306
		220	696	10094	709	938	471	4223	308
1951-52	1938	3494	4543	2597	1492	3214	3425	17378	834
		525	782	1729	1168	1135	1009	11657	322
		492	817	1738	1204	970	1088	11267	372
		492	815	1739	1205	980	1094	11228	375
	1948	3494	4543	2597	1492	3214	3425	17378	834
		291	1803	6469	335	319	290	2221	492
		306	1661	6079	466	358	249	2235	474
		308	1633	6044	461	363	244	2241	467
1959-1960	1938	3464	3707	3775	1805	3251	3892	11919	684
		818	968	1543	1126	827	1611	8261	275
		831	882	1402	1179	767	1568	8171	301
		830	887	1407	1177	769	1567	8169	302
	1948	3464	3707	3775	1805	3251	3892	11919	684
		828	2786	6993	503	532	587	822	1457
		736	2497	6371	886	782	495	942	1338
		705	2386	6276	837	809	497	928	1286
	1951-52	3464	3707	3775	1805	3251	3892	11919	684
		165	318	314	165	154	248	1421	249
		122	401	366	149	135	249	1146	276
		117	405	370	151	135	245	1160	278

Note: All values are in 10^{-4} bits

wellknown, but it is interesting to see that it can be measured quantitatively. There are no other diagonal values (6.1) which are larger than 2. This explains the fact that the Communist figures of Table 3 are mostly so much larger than the figures of the other regions.

Another interesting case is that of Germany (region 3) and Other E.E.C. countries (region 4). Here there are three nonzero flows involved: x_{34}, x_{43}, and x_{44}. The development of the corresponding logarithms (6.1) is as follows:

	(3, 4)	(4, 3)	(4, 4)
1938	.885	.654	.431
1948	1.811	.316	.417
1951-52	1.160	.802	.461
1959-60	1.119	1.079	.634

The transition from 1938 to 1948 is mainly dominated by the war effect. During the fifties regions 3 and 4 became increasingly more important customers of region 4. In the last few years of the period considered here this may have been caused by the establishment of the European Economic Community. A further analysis, based on more extensive data, must decide on the question to what extent this indeed is the case.

REFERENCES

[1] W.E. Deming, Statistical Adjustment of Data. New York, 1944 (Second Printing).

[2] L.S. Schwartz, Principles of Coding, Filtering, and Information Theory. Baltimore-London 1963.

[3] R. Stone and A. Brown, A Computable Model of Economic Growth, Volume 1 of the series A Programme for Growth. London 1962.

[4] C.B. Tilanus and H. Theil, "The Information Approach to the Evaluation of Input-Output Forecasts." Report 6409 of the Econometric Institute of the Netherlands School of Economics (1964). To be published in Econometrica.

[5] J. Waelbroeck, "Une nouvelle méthode d'analyse des matrices d'échanges internationaux." Cahiers économiques de Bruxelles, No. 21 (1964), pp. 93-114.

APPENDIX

As an example of the observed and predicted fractions y_{ij}, $\hat{y}_{ij}$ we present in Table 4 the 1959-60 values, below these the $\hat{y}^3_{ij}$ forecasts derived from the 1951-52 table, and at the bottom the $\hat{y}^3_{ij}$ forecasts derived from the 1938 table. All figures are to be divided by 10^{-4}.

TABLE 4. OBSERVED AND PREDICTED TRANSACTION SHARES, 1959-60

Region of origin	Region of destination 1	2	3	4	5	6	7	8	Total
	Observed transaction shares								
1	569	301	87	184	167	74	15	616	2012
2	312	58	69	55	60	21	17	104	696
3	85	60	-	250	37	167	40	237	877
4	130	67	226	284	95	195	91	306	1394
5	129	38	34	79	-	84	23	410	797
6	61	34	116	107	91	91	38	107	645
7	8	15	35	48	35	40	763	262	1205
8	373	62	202	306	406	71	206	748	2374
Total	1667	635	768	1314	891	743	1193	2789	
	Forecasts $\hat{y}^3_{ij}$ derived from the 1951-52 table								
1	586	319	139	159	138	70	0	601	2012
2	359	45	54	62	47	25	4	101	696
3	71	75	-	260	46	211	21	191	877
4	110	66	196	239	119	155	40	469	1394
5	98	40	25	74	-	87	9	463	797
6	55	39	125	110	94	74	53	95	645
7	17	3	24	38	34	48	832	209	1205
8	371	48	205	371	413	72	234	661	2374
Total	1667	635	768	1314	891	743	1193	2789	
	Forecasts $\hat{y}^3_{ij}$ derived from the 1938 table								
1	531	269	77	166	233	86	174	476	2012
2	301	44	92	48	63	24	19	105	696
3	36	84	-	187	27	132	203	209	877
4	152	70	1690	256	116	162	76	393	1394
5	145	33	26	59	-	69	66	398	797
6	49	51	77	176	82	68	85	57	645
7	93	51	133	178	75	155	228	290	1205
8	361	33	193	243	294	45	342	862	2374
Total	1667	635	768	1314	891	743	1193	2789	