الجمهورية العربية المتحدة

معهد التخطيط القومى

Memo. N° 606

A Multiple-Regression Analysis of International Trade Flows - Its Setup, and Main Results.

By

Hans Linnemann

and

Mary N. Youssef

Operations Research Centre.

القاهرة

٣ شارع محمد مظهر - بالزمالك

A Multiple-Regression Analysis of International Trade Flows - Its Setup, Computer Programmes and Main Results

By

Hans Linnemann and Mary N. Youssef

Contents

1. Introduction. The scope of the report.

Since the late forties, problems of the economic development of the materially less-advanced countries have received a good deal of attention from economists all over the world. Economic development has been re-discovered as both an academical and a practical subject of paramount importance. This study of development problems has had a stimulating influence also on several related fields of study in the economic discipline. A clear example of such a refreshing and stimulating effect of development questions on other areas is the increased attention to international trade which has demonstrated itself during the last decade or so. A substantial number of publications devoted to international trade matters - particularly in the context of the "developing world" - bears witness of a revived interest in this traditional field of economic study.

Most of the work done in this field is, however, of a qualitative rather than a quantitative nature. Some of the more recent studies analyze with a good deal of refinement and detail how economic growth and development might affect the future growth of world trade, or vice versa how the development of a country's or a region's exports might determine its rate of economic growth. Such analysis is usually concerned with marginal relationships, and not with average or total magnitudes. In contrast to these general tendencies, the present study is predominantly quantitative in nature, and it concentrates on average rather than on marginal economic relations.

The approach followed in this analysis originates from a similar but less ambitious study undertaken by a research team of the Netherlands Economic Institute; this research is reported on in an appendix to Professor Tinbergen's recent book on international economic policy[1]. The main objective of the present study is to explain, in a quantitative way, the value

---

1) Jan Tinbergen, Shaping the World Economy-Suggestions for an International Economic Policy, New York: Twentieth Century Fund, 1962.

of the trade flow between any pair of countries in the world. What factors determine the amount (in value terms) that country A exports to country B, and what is the relative importance of the various factors? And once the role played by the basic factors is established, which countries show the most pronounced deviations from the normal pattern of trade flows, and why ?

It is obvious that an adequate discussion of all theoretical and practical questions involved in this study surpasses the scope and size of a simple memorandum. A more elaborate treatment of these questions will be given in a forthcoming publication. The present memorandum concentrates therefore on the empirical part of the study, it describes the computational procedure and the main results obtained. The first author, who has been working at the Institute of National Planning for two years as a visiting expert on economic planning, has initiated the study and is primarily responsible for its set up and for the economic and statistical aspects of the analysis. The second author, who is permanently on the staff of the Operations Research Center of the Institute, is responsible for the computer programmes and computational procedure in general.

The authors are greatly indebted to the Board of Directors of the Institute of National Planning for their permission to use the IBM 1620 computer of the Institute for the purpose of the present study. Their special thanks are due to Dr. Salah Hamid, Director of the Operations Research Center of the Institute, for his generous cooperation and assistance and for stimulating support of their work. Unnecessary to say that, without this support and cooperation from the side of the Institute's management and staff, this study could not have been undertaken and completed. For errors and shortcomings, however, the authors alone are to be held responsible.

2. Factors influencing the size of international trade flows.

In the subsequent analysis, four different magnitudes, or "variables", are used in explaining quantitatively the size of the annual flow of trade between countries. These magnitudes are

(i) national income, or gross national product;
(ii) population size;
(iii) geographical distance;
(iv) political economic preferences.

The latter concept is, strictly speaking, not a measurable magnitude; we will see later on how such a non-quantitative concept can be introduced in a quantitative analysis. First, we will explain briefly the economic significance of the above-mentioned factors.

The economic magnitude or variable that we want to "explain" statistically is the size of the trade flow, measured for an annual period, between any two countries of the world. The most obvious factor influencing a trade flow is the economic size of the trade partners. The United States of America is the world's biggest economic unit as far as the production of goods and services is concerned. It is not surprising, therefore that for many countries the United States is the main supplier of foreign goods. The size of the national product of the exporting country is clearly one of the forces that plays a part in explaining the size of a trade flow. On the other hand, the national product (or the national income) of the importing country is relevant as well. The exports of the United States to Italy are bigger in size than those going to the U.A.R., because the buying power of Italy is larger than that of the U.A.R. Similarly, the United States exports

to the U.A.R. are bigger than those to Libya because Libya's national product or income is much smaller than that of the U.A.R. Thus, we find that the G.N.P.'s (gross national products) of both the exporting and the importing country affect considerable the volume of trade between two countries.

It is, of course, not only national product which is important. If we compare the two countries Mexico and Sweden, we find that their G.N.P.'s are almost the same; yet Swedish international trade is about twice as high as Mexican foreign trade. This would seem to be due to the fact that Mexico can be more selfsufficient than Sweden (without a loss of efficiency in production), because the opportunities for large-scale production are easier realized in a country of 35 million people than in a national economy comprising only 8 million consumers. An analogous comparison could be made between Turkey and Denmark, having about the same total income but substantially different imports and exports. We may note in passing that equal G.N.P.'s with different population sizes also implies a difference in per capita incomes. To the extent that the relative importance of foreign trade would also depend on the level of income per head of the population[2], it is again through the incorporation of population size as a trade - determining variable that this factor is taken care of. Therefore, the population-size variable brings into the analysis two elements: (a) the existence of economies of large-scale production, or at least the idea of indivisibilities leading to a minimum output level for efficient production, and (b) the possible effect of per capita income on foreign trade.

The magnitudes discussed so far are, in a sense, the general characteristics of a country as a supplier or a buyer

2) There is little empirical evidence on the impact of per capita income on the relative importance of foreign trade.

on the world market. Thus, the size of the trade flow from country A to country B depends on the general characteristics of A as a supplier to the world market (i.e. the G.N.P. and the population of country A) and the basic variables describing B as a buyer on the world market (i.e. again the G.N.P. and the population size, but now of course those of country B). However, trade between two countries also depends on specific factors, that is, on factors which take a particular value in each individual case of one pair of trade partners. Two such factors are introduced in the analysis; the first of them is the geographical distance between the trade partners.

The distance between country A and country B influences the trade between these nations in a negative way: the greater the distance, the smaller the amount of goods traded between the partners. This negative influence is due to two causes. First of all, and most obviously, the greater the distance between the trade partners, the higher the cost of transportation and hence the weaker the competitive position of the exporting country. Secondly, even in our present-day world with its tremendously improved possibilities of communication, exporters are still much better informed about business conditions and the market situation in neighbouring or nearby countries than about commercial conditions and economic possibilities at the other end of the globe. The foreign trade statistics reveal that usually a country's neighbours are among its most important trade partners. The impact of the factor distance on foreign trade flows is clearly discernible.

The second specific factor affecting the trade between a pair of countries is the degree of political and socio-economic affinity that exists between the two partners. Close political cooperation, and a thorough knowledge of each other's culture,

language, and institutional setup, will have a stimulating affect on trade. This is particularly true if the political-economic cooperation expresses itself in the establishment of preferential trade areas, i.e. in the lowering or abolition of import duties on each other's commodities, and the partial or total remoral of other artificial impediments to mutual trade.

The present study starts from the simplifying assumption that, with a number of important exceptions to be specified below, the degree of socio-economic affinity is roughly the same for all countries of the world. There may be positive or negative deviations in individual cases from this "average" level of cooperation-implying also an "average" level of trade impediments like import duties and the like-but these deviations are considered to be random deviations from the statistical viewpoint. In making this assumption, the necessary exceptions immediately spring to the eye. First of all, the trade with the centrally-planned economics of Eastern Europe and Asia does not fit into this assumed normal pattern or level of cooperation. In spite of the recent trend towards increased East-West trade, there are still numerous embargoes and important restrictions on the trade between communist and non-communist countries. In first instance, this problem could only be solved by excluding all communist countries from the analysis. Yugoslavia has been included, however, as this country has pursued for many years already a more liberal foreign trade policy. A second important exception is due to the Arab League embargo on trade with Israel. The trade flows between Israel, on the one hand, and the nine Arab League member countries included in the analysis, on the other, were zero for political reasons and have, therefore, been excluded in the empirical analysis. For the year to which the empirical investigation refers, 1959, no other boycotts or trade embargoes were effective as yet (Cuba, Portugal, South Africa).

Having thus excluded from our analysis the trade flows for which the negative deviations from our assumption on "normal" trade relations are clearly due to systematic forces, and too pronounced to be considered as a random disturbance, we have to deal now with systematic positive deviations. These positive deviations are found to exist in the preferential trade agreements originating from past- and sometimes still present-colonial relations. Trade between the member countries of the British commonwealth is fostered by the so-called " Imperial preferences" (and, of course, by the greater familiarity with each other's economic and political conditions). Similarly, members of the French Community have much easier access to each other's markets than third countries have. Finally, in 1959 the Belgian and the Portugese colonial "empires", though dwindling, were still in existence. In all these cases, trade between member countries was much larger than usual, as the following sections will show. The preferential trade relations were taken into account in the empirical analysis by introducing a so-called "dummy variable". This variable has the value zero if no preferential trade relations exist between the pair of countries under consideration; if both countries belong to the same preferntial grouping, the variable takes a value (an arbitrarily fixed positive value, say 1) different from zero. For 1959, the trade-fostering effects of more recently established preferential areas, like the European common Market, or the latin-American Free Trade Association, were still too insignificant to warrant the introduction of additional dummy variables.

Summarizing the above discussion, we see that the following variables contribute to an explanation of the size of trade flows. We introduce symbols for the variables, and denote the exporting country's symbols with a subscript i

and the importing country's with a j; a double subscript refers to a flow from i to j. The variable to be explained:

X_{ij} = the annual trade flow from i to j.

The explanatory variables:

Y_i = gross national product of country i ;

N_i = population of country i ;

Y_j = gross national product of country j ;

D_{ij} = distance between country i and country j ;

P_{ij} = preferential relationship between country i and country j.

The last-mentioned variable can be "split up" further. This is desirable as it is not certain that British commonwealth preference, French Community preference, and the Belgian and portugese colonial preferences have all of them an equally strong trade-fostering effect. Thus, actually three preferential variables are distinguished in the empirical exercise:

P^B = British Commonwealth preference;

P^F = French Community preference ;

P^C = Belgian and Portugese colonial preferences.

In specifying the mathematical form of the relationship we adhered to the well-known principle that demand and supply relationships are often of a multiplicative rather than additive character. The relation between the variables is supposed to be not linear in the absolute values but instead logarithmic-linear. Denoting the unknown parameters of the relationship by the Greek letter φ, we have

$$X_{ij} = \varphi_o Y_i^{\varphi_1} N_i^{\varphi_2} Y_j^{\varphi_3} N_j^{\varphi_4} D_{ij}^{\varphi_5} (P_{ij}^B)^{\varphi_6} (P_{ij}^F)^{\varphi_7} (P_{ij}^C)^{\varphi_8}$$

After the foregoing discussion it will be clear that the following values of the nine parameters are to be expected:

$o', 2', 3', 6', 7', 8'$: Positive

$2', 4', 5$: Negative

The constant $_o$ is merely a scale factor; its value depends on the choice of the units of measurement.

The above formula will now be confronted with actual data, in order to determine (a) how satisfactory or unsatisfactory the assumed relationship explains actual trade; (b) what values the parameters take; and (c) what other systematic forces influencing the trade flows could be traced through studying the "residuals", i.e. through observing the non-explained variations in trade flows. The technique to be used in this empirical study is that of ordinary least-squares multiple regression analysis. Before we go more deeply into technical details, however, we have to describe the geographical coverage of the study and to mention the sources of the necessary data.

3. Coverage of the empirical study, and sources of data.

As the trade-flow formula presented in the previous section is intended to be of general applicability, it is necessary to verify its accuracy and "prediction power" on the basis of a set of data which is as broad and encompassing as possible. The geographical coverage has been made therefore as wide as the data permitted. Apart from the Communist states, all countries of the world have been included except (a) those for which no data

are available, or for which the trade data are apparently highly unreliable; and (b) those countries for which most of the foreign trade takes the form of transit trade (Hong Kong, Singapore, Panama, etc.). Because of the first requirement- more or less accurate data available- a rather large number of countries dropped out. We are left with 80 countries altogether, for our empirical analysis. These countries- or groups of countries, as for instance in the case of Former French West Africa, comprising eight countries nowadays independent- are listed alphabetically in Table 3.1; their code number corresponds with the ranking on a regional basis which is used in the trade statistics.

Table 301

The eighty countries included in the empirical analysis
(the code number indicated geographical ordering)

65	Afghanistan	27	France	80	New Zealand
53	Angola	28	Germany, Fed. Rep.	10	Nicaragua
13	Argentina	58	Ghana	62	Nigeria
79	Australia	29	Greece	34	Norway
23	Austria	07	Guatemala	70	Pakistan
24	Belgium	08	Honduras	17	Paraguay
14	Brazil	30	Iceland	18	Peru
21	British Guiana	69	India	77	Philippines
66	Burma	73	Indonesia	35	Portugal
71	Cambodia	46	Iran	63	Rhodesia
55	Cameroun	47	Iraq	52	South Africa
02	Canada	31	Ireland	36	Spain
67	Ceylon	49	Israel	51	Sudan
15	Chile	32	Italy	37	Sweden
16	Columbia	11	Jamaica	38	Switzerland
54	Congo (Léop.)	74	Japan	43	Syria
03	Costa Rica	48	Jordan	68	Taiwan
04	Cuba	75	Korea, Rep. of	78	Thailand
41	Cyprus	44	Lebanon	12	Trinidad
25	Denmark	50	Libya	64	Tunisia
05	Dominican Rep.	59	Madagascar	39	Turkey
06	El Salvador	76	Malaya, Fed. of	42	United Arab Rep.
45	Ethiopia	09	Mexico	22	United Kingdom
26	Finland	60	Morocco	01	United States
57	FFr. Equat. Africa	61	Mozambique	19	Uruguay
56	FFr. West Africa	33	Netherlands	20	Venezuela
				72	Vietnam, Rep. of
				40	Yugoslavia

Note: the following abbreviations are used for simplicity.

FFr. Equat. Africa: the countries formerly constituting French Equatorial Africa.

FFr. West Africa: the countries formerly constituting French West Africa.

Rhodesia: the Federation of Rhodesia and Nyasaland.

South Africa: South Africa and South West Africa together.

Trinidad: Trinidad and Tobago.

For these eighty countries, the trade flows expressed in United States dollars can be obtained directly from the publication Direction of International Trade, a joint publication of the United Nations, the International Monetary Fund, and the International Bank for Reconstruction and Development. This gives us 80 times 79 = 6320 trade flows. Some of these observations have to be dropped, however, because of the Arab League embargo on trade with Israel. Nine Arab countries included in our study are involved in this boycott: Iraq, Jordan, Lebanon, Libya, Morocco, Sudan, Syria, Tunisia, and the United Arab Republic. Thus, 2 times 9 = 18 trade flows are zero for political reasons. Moreover, because of the merger between Egypt and Syria, no figures for the trade flows between them in 1958 and 1959 are published. For these reasons 20 trade flows have to be omitted from the analysis; our study is based on the remaining 6300 observations of the variable to be explained, the trade flow between a pair of countries. Together these 6300 trade flows account for about 83 percent of total world trade in commodities, excluding trade with and within the Communist bloc.

In order to reduce the effect of incidental transactions of unusual size, and of incidental difficulties in trade contacts, all trade flows have been taken as three-year averages for 1958-1960, rather than as direct observations for 1959. On the whole, the trade-flow values were not greatly affected by this procedure, but certain irregularities could be eliminated or at least moderated in this way.

It is an inevitable shortcoming of this approach that the trade flows studied include only merchandise trade. There are data on trade in services, but for a more limited number of countries and, more essential, for the national aggregates only.

Thus, total imports or exports of services are known for at least the more important countries, but the countries of origin or destination of the services are not given in the statistics.

It should be noted that the size of a trade flow can be measured either at the point of export, or at the point of import. Apart from differences in valuation exports are usually valued on a free-on board basis, and imports at the higher cost-insurance-freight prices and some other minor distortions, these two measurements should give identical results. In practice, the differences turn out to be rather substantial sometimes. In the empirical analysis we use two sets of trade-flow data, one obtained from export statistics, and another giving the data from the import statistics. We may use the symbols X^{E}_{ij} and X^{M}_{ij}, respectively, to distinguish between the trade flow figures from the two sources.

Concerning the explanatory variables, the most difficult and time-consuming collecting of data was that with respect to the distance variable. Distances have been measured as the shortest navigable distance between the main ports of the respective countries, plus the overland distances from the ports to the economic "gravity points" of the countries concerned. Overland distances have been added directly to sea distances, without applying any correction factor for higher overland costs. Sea distances were calculated from S. Luensee, Entfernungstabellen, (3 Vols.), Hamburg, 1960/1961

Population figures were obtained from the well-known United Nations sources. Gross national product figures are those compiled by P. Rosenstein-Rodan in "International aid for underdeveloped countries", Review of Economics and Statistics, May 1961. We used alternatively nominal G.N.P. figures, and

real G.N.P. estimates. The latter figures purport to give an indication of a country's G.N.P. when measured at American prices. Except for Canada and Venezuela, and of course for United States itself, the real G.N.P. figures are higher than the nominal G.N.P. values.

Table 3.2

Groups of countries having preferential trade relations with each other

1. The "British Commonwealth of Nations" -group.

02	Canada	62	Nigeria
11	Jamaica	63	Rhodesia
12	Trinidad	67	Ceylon
21	British Guiana	69	India
22	United Kingdom	70	Pakistan
31	Ireland	76	Malaya
41	Cyprus	79	Australia
52	South Africa	80	New Zealand
58	Ghana		

2. The "French Community" - group.

27	France	60	Morocco
55	Cameroun	64	Tunisia
56	FFr.West Africa	71	Cambodia
57	FFr.Equat.Africa	72	Vietnam
59	Madagascar		

3. Other preferential groups.

3 A.	35	Portugal	3 B.	24 Belgium
	53	Angola		54 Congo (Leop.)
	61	Mozambique		

Note: The countries listed as members of group 1 and group 2 are not necessarily members of these groupings from a legal-political point of view.

The preference variable has been given the values 1 or 2, depending on the fact whether preferences did (value 2) or did not (value 1) exist. As the calculation is made in logarithms, the case of no preferences is characterized by zero (log 1=0). Table 3.2 lists the countries included in one of the three preferential trade areas that are distinguished.

The units of measurement of the variables discussed in this section are as follows;

X_{ij} : Millions of U.S. dollars

Y_i, Y_j : Millions of U.S. dollars

N_i, N_j : Millions of inhabitants

D_{ij} : Thousands of nautical miles (1 nautical mile = 1,852 meters).

P^B_{ij}, P^F_{ij}, P^C_{ij} : take the value 2 if preferential treatment exists between i and j; otherwise the value is 1.

4. Statistical aspects of the empirical analysis.

As has been mentioned already, the statistical technique used for testing the empirical usefulness of the trade flow equation and for determining the numerical values of the parameters is the ordinary least-squares regression analysis. Applying least-squares regression to the logarithms of the trade flow values implies that not the obsolute residuals are minimized, but instead the relative (say, the percentage-wise) residuals. In view of the fact that the obsolute values of the original observations vary rather widely and have a very

uneven frequency distribution-a great concentration in the low-value intervals-, the necessary condition of homoscedasticity of the residuals is more likely to be fulfilled when working with logarithms than when applying regression analysis to the original values in antilogs. This is, therefore, another argument in favour of a logarithmic analysis.

In another respect, working with logarithms confronts us with a special problem, however. For a surprisingly large part of our basis set of 6300 trade flows (actually for 40 to 45 percent of the total number of observations), annual trade is zero or at least too low to be recorded in the statistics. This is illustrated in the frequency distribution shown in table 4.1 If we would neglect these zero observations, we would eliminate a substantial and in all respects essential part of our sample; this is therefore absolutely unacceptable. But if we want to include them, we face the difficulty that taking the logarithm of zero is mathematically an undefined operation and thus impossible. The usual procedure in such cases is to put an arbitrary but small positive value instead of zero. In the present case, however, this is not an easy matter, for the following two reasons.

Table 4.1

Frequency distribution of the size of trade flows between 80 countries (average of 1958/1960; millions of U.S.Dollars).

Reported size of the trade flow		Export statistics	Import statistics
Nil, i.e. <	0.1	2900	2768
	0.1	386	374
0.2 - <	1.0	913	954
1.0 - <	10.0	1235	1299
10.0 - <	100.0	698	723
100.0 - <	1000.0	163	175
1000.0 or greater		5	7
Total number		6300	6300

To start with, it should be realized that the trade flow reported zero in the statistics may in actual fact well be different from zero, as the smallest unit of measurement in the trade statistics concerned is $ 100,000. This means that an annual trade flow of, say, $ 35,000 will not be recorded (flows of more than $ 50,000 will be rounded to $ 0.1 million, the smallest unit).

So, we would have to assume an arbitrary figure of less than $ 50,000 for the trade flows reported zero, if we want to include such flows. In a linear analysis, it would not matter very much which figure below 0.05 million dollars we would take. But it makes in logarithms quite a difference whether we put this assumed at 0.04 million dollar, at 0.01 million, or at 0.001 million, and so on. The natural logarithms of these figures (the computer works in natural logarithms) are -3.219, -4.605, and -6.908, respectively, which values are by no means

"near" to each other. (Note that the largest observation in the set of data has a logarithm of + 8.320).

The second difficulty resides in the fact that the "zero flows" constitute such a large part of the total number of observations. Therefore, the assumption made about the small value to be substituted for the zero will probably affect the final result rather considerably. We have to be careful in choosing the method by which we incorporate the "zero flows" in the regression analysis.

In view of the above-mentioned problems, the following procedure has been adopted. In first instance, all "zero flows" are simply neglected and excluded from the analysis. With the parameters estimated in this first round, we calculate the "expected" or "explained" values of all trade flows. If the "explained" value of a zero flow turns out to be lower than $ 50,000 - i.e. "zero", in terms of the units of measurement of the trade statistics -, then apparently the estimated scope and level of the regression line are not in conflict with the fact that the trade flow in question is recorded as being zero. The parameter estimates would not have been different if the "zero flow" had been taken into consideration explicitly, in the correlation. However, if the "explained" value of the zero flow is bigger than $ 50,000, the estimated parameter values are not able to account for the actual fact of its being statistically recorded as zero; they are therefore in all probability biased in an "upward" direction. Consequently, these unexplained cases may not be neglected in the analysis, and have to be introduced by giving them some arbitrary value lower than $ 50,000. A second round of calculations is necessary. This iteration process could be continued for still further rounds, along the same lines.

The advantage of this procedure is that it reduces the possible errors due to the choice of this arbitrary value below 0.05 million, because it limits the number of cases in which such an assumed value has to be introduced. Moreover, the assumed value could now be chosen relatively close to the limiting value of $ 50,000, as the really low values (close to zero indeed) will have been eliminated in the first round already. In the present analysis, two different assumptions were made with respect to these "zero flows" in the second round of calculations, namely $ 0.01 million and $ 0.02 million, respectively. In this second round, the number of observations included was increased to 4,831, as compared with 6300-2900 = 3,400 and 6300 - 2768 = 3,532 in the first round of calculations (for the latter two figures, see Table 4.1).

The information that we sought to obtain through the regression analysis comprised the following numerical magnitudes:

(i) an estimate of the parameters of the trade flow equation (φ_0 to φ_5);

(ii) the standard errors of the parameter estimates;

(iii) the coefficient of determination (R^2) and the regression coefficient (R);

(iv) the variance and standard deviation of the residuals;

(v) the parameter estimates when all variables are measured in terms of their standard deviations. These so-called beta coefficients indicate the relative importance of the individual explanatory variables in explaining the variations in the dependent variable, the trade flows;

(vi) the "explained" values of the trade flows;

(vii) the deviations between observed and "explained" trade flow values, both in absolute terms and percentage wise.

The above results were obtained for a fairly large number of different cases. Let us list these cases with a short description of their characteristics.

(1) X^E explained by the eight variables introduced before, G.N.P. in nominal terms (Y_i^N and Y_j^N); number of observations 3,400;

(2) X^E explained by ditto, but G.N.P. in real terms (Y_i^R and Y_j^R); number of observations 3,400;

(3) X^M explained by ditto, with Y^N; number of observations 3,532;

(4) X^M explained by ditto, with Y^R; number of observations 3,532;

(5)-(8) the same as cases (1)-(4), except that those observations where X^E and X^M differed very considerably (one more than twice the value of the other) were rejected, because of their apparent unreliability and inaccuracy; number of observations 2,740;

(9)-(16) the same as cases (1)-(8), but with a restriction on the estimated parameters; on a priористic grounds it might be assumed that -under certain conditions- the G.N.P. and the population parameters should have the same value at the export side and at the import side of the trade flow equation, i.e. $\varphi_1 = \varphi_3$ and $\varphi_2 = \varphi_4$ (this restriction would correspond with bilaterally balanced trade, in economic terminology). In these cases we apply so-called conditional regression analysis; number of observations: the same as in the original cases (1)-(8);

These first sixteen cases together form the first round of calculations. On the basis of the results, part of the "zero-flows", so far excluded, were incorporated in the

analysis. For the "zero" X^E observations, an arbitrary value of 0.01 million dollars was assumed; for the "zero" X^M observations, this value was set at 0.02 million dollars. The second round thus consisted of the following cases:

(17)-(20) the same as (1)-(4) above, but including now the selected "zero flows"; number of observations 4,832;

(21)-(24) the same as cases (17)-(20), but with a priori condition that $\varphi_1 = \varphi_3$ and $\varphi_2 = \varphi_4$; number of observations again 4,832.

Finally, it should be noted that the same analysis was applied to the trade of individual countries. For each exporting country individually, the parameters φ_3 to φ_8 may be estimated (although at least two of the three preferential-trade parameters will have to be zero, as a country can belong to one preferential trade area only). The parameters φ_1 and φ_2 cannot be estimated, obviously, as the corresponding variables always refer to the same country and therefore do not take different values. Similarly, six parameters can be estimated for importing countries individually (not φ_3 and φ_4 in this case). For individual exporting countries, these country-by-country estimates were made in cases (1), (3) and (4); for individual importing countries, a country-by-country analysis was made for cases (17)-(20). We may denote these cases by (1A), (3A), etc.

The treatment of preferential trade relations differed somewhat in the second round of calculations, as compared with the original setup. P^B still refers to British Commonwealth trade, but it excludes now the trade with the United Kingdom itself and it comprises only the preferential relations between all non-U.K. Commonwealth partners. Similarly, P^F refers to the French Community trade except the trade with France itself. P^C covers, in its new meaning, only the trade flows from and

to the "metropolitan country" of the preferential grouping of nations. P^C thus refers to, for example, U.K.-Ghana, India-U.K., France-Morocco, Madagascar-France, Congo (Leop)-Belgium, Portugal-Angola, etc. It does not refer to Ghana-India (which is P^B) or to Morocco-Madagascar (which is P^F). The reason for this change is that the trade-fostering effects of preferential trade areas are much more pronounced in the relations with the (colonial) "mother country" than in the relations between the junior members of the ex-colonial trading areas. This can be seen very clearly from the results of the analysis given below.

5. The computational set-up of the analysis.

The multiple regression analysis described in the preceding sections has been performed on the IBM 1620 electronic computer, of the Operations Research Center of the Institute of National Planning, Cairo. The computer programmes were formulated in FORTRAN 2, and are listed in the appendix. In the present section we will give a brief survey of the various steps involved in the correlation analysis.

The first step to be taken was, obviously, to put the set of basic data on punch cards. All variables with a double subscript (ij) are specific variables, i.e. we have 6,300 different observations. They had to be punched therefore in an equally large number of cards. The variables with one subscript only (whether i or j) take only 80 different values; they could be punched separately in a set of 80 cards only, and might be called general variables. The first computer programme was a simple one, namely to transform these sets of data from anti-logs into logarithms. As was remarked already, these are natural or Napierian logarithms, with e as base.

As is usually done in a multiple regression analysis- particularly when it is a large-size computation, as in the present case- the required product-sums of the variables are not calculated directly in terms of the deviations of the variables from their mean value. Instead, the product sums of the absolute values were calculated first, and these sums were corrected afterwards for not having measured the variables as deviations from the mean. This well-known procedure implies that the product sums of absolute values have to be reduced by the product of the means of the variables concerned multiplied by the number of observations. The second computer programme thus calculated the sums (as the basis for the means) of the variables, and the third programme computed the product sums for the absolute values and-after subtracting the result of the second programme-the product sums for the deviations from the means.

A fourth computer programme was designed for inverting the matrix of product sums. Following the inversion, the same programme calculated the regression coefficients, the coefficient of determination and the variance of the residuals (in logs) as well as their square roots, the standard errors of the regression coefficients, and the beta coefficients. Because of the conditional regression to be applied, the programme was formulated in such a way that both a 9x9 matrix (the normal case) and a 7x7 matrix (the conditional case) could be inverted and further processed.

Programme number 5 computes the "explained values" in terms of logarithms, by applying the regression coefficients obtained in the previous phase to the basic data for the explanatory variables. It was explained in the previous section that the analysis covered by the first four computer programmes has been made for a number of cases. In the first round of calculation, 16 cases were computed, and therefore 16 sets of regression

coefficients were obtained. From these 16 sets , seven sets were selected as data for the fifth computer programme. This programme gave us therefore seven different "explained values", still expressed in terms of natural logarithms.

The sixth and final programme transformed the "explained values" from logs into antilogs. The antilogs were compared with the original data, and the positive or negative residuals or deviations in normal values were fixed. In two cases (out of the six or seven computed) also the percentage-wise deviations were determined, by dividing the deviation through the original observation. The deviations were also summed, both over the exporting and over the importing country concerned, and divided by the total exports and total imports, respectively, in order to determine the relative size of the sum of the deviations. The export and import totals were obtained in this programme as well.

7. Concluding remarks.

The reader who has gone through the preceding sections will realize that in many respects more could have been said about the subject matter than what is actually done in this text. The present memo only mentions the basic setup of the analysis, and it reports the main results without much comment. A more comprehensive and elaborate discussion will be given in a forthcoming publication, as soon as all computations are finished. This document is essentially an interim report on the progress of the work, and not a final report. The following remarks similarly are of a rough and tentative character; they may have to be modified later on in the light of the final results.

Our first remark is that all explanatory variables introduced in the analysis have shown themselves to be statistically

significant factorcs. There can be no doubt that all of them contribute to a statistical explanation of world trade flows. Yet the total "explanation" is far from perfect. Quite often, individual trade flows reveal discrepancies between explained and actual trade which imply a proportionality factor of 2 to 4. Thus there is still much scope for improvement. In particular, the commodity composition of total foreign trade of the exporting and the importing country will be introduced in further studies as an explanatory variable.

Second, the parameter estimates obtained in the second roun of calculations-cases (17) to (24) - differ rather significantly from those obtained in the first round. A frequency distribution of the relative deviations of explained from actual trade flows in the two cases reveals that for smaller trade flows the second-round parameters lead to the best explanation, whereas the bigger trade flows are better explained when using the parameters of the first round. This suggests that the relationship is not logarithmic-linear. The non-linear relation may be approximated by two linear relationships; the "breakeven point" (the point at which the two sets of parameters give about equally "good results) corresponds with a trade flow size of $ 1.0 million. Trade flows smaller than $ 1.0 million are explained by the second-round relationship; flows of $ 1.0 million or more follow more closely the pattern implied by the first-round parameters.

Third, we observe that in all cases the explanatory variable "real G.N.P." leads to a somewhat better result (i.e. a somewhat higher correlation coefficient) than "nominal G.N.P.", but the difference is very small indeed. The results also show that there is some intercorrelation between the G.N.P. variable and the population size; this is, of course, not very surprising. This intercorrelation makes the estimates of β_1 and β_3 on the one hand

and of $_2$ and $_4$ on the other, less reliable. The treatment of preferential trade relations is more satisfactory in the second round than in the first one, as can i.a. be seen from the standard errors of the estimated parameters. This can be verified also form the results of the country-by-country parameter estimates but at the time of writing these results were not yet avavilable.

Fourth, it is interesting to see what results are obtained when the individual trade flows are summed for either the exporting or the importing country. Comparing actual and explained trade after summation, we find that the relative deviations are now much smaller (as positive and negative deviations may partly offset each other). When we make this summation for cases (1) and (4) - two cases which we have used most frequently for more detailed analysis -, we find that explained total trade differs in both cases (1) and (4) less than 10 per cent from total actual trade for the following seven countries: Portugal, Switzerland, Cyprus, the United Arab Republic, Ethiopia, Israel, and Libya (according to the first-round calculations). For the U.A.R., the deviations are only 7.3 per cent and 3.6 per cent, respectively. Thus, the U.A.R. fits into the general explained pattern quite well. Some countries persistently show negative deviations, both at the export and at the import side for all cases: France, Iceland, Spain, Ethiopia (for the latter country, the negative deviations are very small only). This means that their trade with other countries is obviously less than usual. Other countries have a normal import situation, but a (strongly) negative export deviation; in this group we find the countries heavily dependent on foreign assistance (Jordan, South Korea), or having an abnormally high share of services exports (Greece). These and similar results can easily be interpreted as to their economic meaning.

Fifth, let us see what the detailed results for the U.A.R. are. As suggested above, we will use the first-round parameters for explaining the larger trade flows, and the second-round parameters for the small values ($ 0.9 million or less). The explained values are compared with the actual ones; the actual figures of the export statistics are compared with the explained values obtained by using nominal G.N.P. figures, whereas the import statistics data are compared with an explanation using real G.N.P. figures. The comparison is made in Tabel. The figures speak for themselves, but a few general comments may be given. The exports to North America are smaller than expected; in the case of the United States this may be attributable to the commodity composition of Egypt's exports (predominantly cotton) - hence the need to incorporate the commodity composition in a future refinement of the analysis. The U.A.R. exports to neighbouring Arab states are greatly underestimated, particularly in the case of the southern neighbour, Sudan. Either the measurement of the distance to these countries has been unsatisfactory, or else the effect of the special ties between the Arab countries is stronger than originally was assumed. Remarkably are also the substantial positive deviations in the trade flows to a number of non-aligned countries with which the U.A.R. closely cooperates in matters of international politics: Yugoslavia, Ceylon, India Indonesia. In the case of Japan, however, it will be the commodity composition of each other's foreign trade which leads to the large positive residual.

We limited ourselves in the above discussion to the exports of the U.A.R. A similar table could be given for the Egyptian imports, and both for exports and imports the results of other cases would be mentioned. All this information is available for each of the 80 countries included in the analysis. Obviously a full discussion of all individual results is hardly possible; in the

present memo we have concentrated on the method of computation rather than on the specific results. A more thorough discussion, both of the techniques of analysis and of the results obtained, will be presented in a forth-coming text of bigger size than the the present memorandum.

Table

Actual and explained trade flows from the United Arab Republic to other countries, average for 1958-1960, in millions U.S. $

	Country of destination	data from export statistics actual	data from export statistics explained	data from import statistics actual	data from import statistics explained
01	United States	14.6	44.5	23.9	46.3
02	Canada	0.4	3.6	0.4	3.8
03	Costa Rica	0.0	0.0	0.0	0.0
04	Cuba	3.1	0.9	0.0	0.2
05	Dominican Rep.	0.0	0.0	0.0	0.1
06	El Salvador	0.0	0.0	0.0	0.0
07	Guatemala	0.0	0.0	0.0	0.0
08	Honduras	0.0	0.0	0.0	0.0
09	Mexico	0.1	0.4	0.3	0.5
10	Nicaragua	0.0	0.0	0.0	0.0
11	Jamaica	0.0	0.0	0.0	0.1
12	Trinidad	0.0	0.0	0.0	0.1
13	Argentina	0.5	0.5	0.0	0.8
14	Brazil	0.3	0.7	0.3	0.8
15	Chile	0.0	0.1	0.0	0.1
16	Colombia	0.0	0.2	0.0	0.2
17	Paraguay	0.0	0.0	0.0	0.0
18	Peru	0.0	0.1	0.0	0.1
19	Uruguay	0.2	0.1	0.6	0.1
20	Venezuela	0.1	0.3	0.1	0.3
21	British Guiana	0.0	0.0	0.0	0.0
22	United Kingdom	16.2	15.3	12.3	17.5
23	Austria	9.0	4.6	3.6	4.9
24	Belgium & Lux.	4.4	4.6	4.7	4.6
25	Denmark	0.8	0.9	1.0	2.6
26	Finland	1.8	1.6	0.9	0.9
27	France	11.0	20.7	11.4	20.9
28	Germany, Fed. Rep.	19.5	13.2	24.5	16.2
29	Greece	2.6	5.0	2.5	5.2
30	Iceland	0.0	0.1	0.0	0.1
31	Ireland	0.1	0.3	0.3	0.5
32	Italy	20.4	16.1	27.9	17.5
33	Netherlands	6.6	4.0	6.4	4.9
34	Norway	0.6	0.7	0.7	1.0
35	Portugal	0.9	0.4	0.8	0.6
36	Spain	11.0	5.0	8.2	5.6
37	Sweden	0.9	1.7	1.1	4.2
38	Switzerland	8.9	6.4	6.1	6.4
39	Turkey	0.7	2.8	1.2	5.3
40	Yugoslavia	20.9	4.1	19.3	4.3

Table .... (continued)

	Country of destination	data from export statistics actual	data from export statistics explained	data from import statistics actual	data from import statistics explained
41	Cyprus	0.1	0.8	0.2	1.2
44	Lebanon	6.0	2.4	8.2	2.1
45	Ethiopia	0.1	0.2	0.2	0.3
46	Iran	0.2	0.2	0.0	0.2
47	Iraq	0.3	0.1	0.5	0.1
48	Jordan	3.1	0.8	2.4	0.6
50	Libya	2.2	0.3	1.6	0.2
51	Sudan	15.3	1.0	12.7	0.8
52	South Africa	1.1	2.0	1.2	2.1
53	Angola	0.0	0.0	0.0	0.0
54	Congo (Léop.)	0.0	0.1	0.0	0.1
55	Cameroun	0.0	0.0	0.0	0.0
56	For.Fr.West Africa	0.0	0.1	0.0	0.2
57	For.Fr.Equat.Africa	0.0	0.0	0.1	0.0
58	Ghana	0.1	0.1	0.1	0.1
59	Madagascar	0.0	0.0	0.0	0.1
60	Morocco	0.3	0.3	0.3	0.3
61	Mozambique	0.0	0.0	0.0	0.0
62	Nigeria	0.0	0.1	0.0	0.1
63	Rhodesia & Nyasal.	0.0	0.1	0.0	0.1
64	Tunisia	0.2	0.2	0.2	0.3
65	Afghanistan	0.0	0.0	0.0	0.1
66	Burma	0.2	0.0	0.4	0.1
67	Ceylon	2.4	0.6	3.5	0.6
68	Taiwan	0.0	0.0	0.0	0.1
69	India	24.0	4.7	21.9	5.9
70	Pakistan	1.4	1.4	0.3	0.3
71	Cambodia	0.0	0.0	0.0	0.0
72	South Vietnam	0.0	0.1	0.2	0.1
73	Indonesia	5.0	1.7	4.9	1.6
74	Japan	18.3	3.8	21.7	4.7
75	Korea, Rep. of	0.0	0.1	0.0	0.1
76	Malaya, Fed. of	0.7	0.2	0.3	0.3
77	Philippines	0.4	0.2	0.4	0.2
78	Thailand	0.0	0.1	0.0	0.1
79	Australia	0.1	0.8	0.2	1.3
80	New Zealand	0.0	0.2	0.0	0.4
	Total export to listed countries	237.1	181.6	240.0	201.5